70 JAAR

1920 1990

tijdschrift van het

nederlands elektronica- en radiogenootschap

deel 55 nr. 2 1990

nederlands elektronica- en radiogenootschap

Nederlands Elektronica- en Radiogenootschap
Postbus 39, 2260 AA Leidschendam. Gironummer 94746
t.n.v. Penningmeester NERG, Leidschendam.

HET GENOOTSCHAP

De vereniging stelt zich ten doel het wetenschappelijk onderzoek op het gebied van de elektronica en de informatietransmissie en -verwerking te bevorderen en de verbreiding en toepassing van de verworven kennis te stimuleren.

Het genootschap is lid van de Convention of National Societies of Electrical Engineers of Western Europe (Eurel).

BESTUUR

LIDMAATSCHAP

Voor lidmaatschap wende men zich tot de secretaris.
Het lidmaatschap staat open voor academisch gegradueerden en hen, wier kennis of ervaring naar het oordeel van het bestuur een vruchtbaar lidmaatschap mogelijk maakt. De contributie bedraagt *f* 60,— per jaar.

Studenten aan universiteiten en hogescholen komen bij gevorderde studie in aanmerking voor een junior-lidmaatschap, waarbij 50% reductie wordt verleend op de contributie. Op aanvraag kan deze reductie ook aan anderen worden verleend.

HET TIJDSCHRIFT

Het tijdschrift verschijnt zesmaal per jaar. Opgenomen worden artikelen op het gebied van de elektronica en van de telecommunicatie.

Auteurs die publicatie van hun wetenschappelijk werk in het tijdschrift wensen, wordt verzocht in een vroeg stadium contact op te nemen met de voorzitter van de redactiecommissie.

De teksten moeten, getypt op door de redactie verstrekte tekstbladen, geheel persklaar voor de offsetdruk worden ingezonden.

De abonnementsprijs van het tijdschrift bedraagt *f* 60,—. Aan leden wordt het tijdschrift kosteloos toegestuurd.

Tarieven en verdere inlichtingen over advertenties worden op aanvrage verstrekt door de voorzitter van de redactiecommissie.

REDACTIECOMMISSIE

ONDERWIJSCOMMISSIE

DIGITALE HDTV

J. Biemond

Technische Universiteit Delft
Faculteit der Elektrotechniek
Vakgroep Informatietheorie
Postbus 5031, 2600 GA Delft

Abstract

An introduction is given into the subject of bandwidth compression of digital High Definition TV (HDTV) for the future Broadband Integrated Services Digital Network (BISDN). The hierarchical structure of the BISDN network together with the call for technological evolution instead of revolution in digital TV suggests a hierarchical coding structure in which both EQTV (a signal according to the CCIR recommendation 601) and the HDTV signal (a signal with twice the spatial resolution of the EQTV signal and an increased aspect ratio of 16:9) coexist. Three possible hierarchical coding structures will be discussed.

I Digitale TV – Waarom?

Naarmate de afmetingen en de lichtintensiteit van het beeldscherm in TV-ontvangers groter worden, worden de beperkingen van de bestaande TV-systemen (NTSC, PAL en SECAM) steeds duidelijker merkbaar en groeit de behoefte aan een nieuw systeem dat aangeduid wordt met "high definition"-televisie (HDTV). Dit moet een vergroting van het scheidend vermogen, een vermindering van de onderlinge beinvloeding van kleur- en helderheidsinformatie, een vergroting van de aspectverhouding en stereofonisch geluid met "hifi"-kwaliteit bieden. Het is hierbij van groot belang dat een compatibel systeem gekozen wordt, zodat bestaande apparatuur bruikbaar blijft en een geleidelijke ontwikkeling (evolutie) van de TV-kwaliteit door de geleidelijke aanschaf van nieuwe apparatuur voor de kijker mogelijk is (Annegarn, 1986).

Het onlangs voor gebruik op geostationaire aardsatellieten (Direct Broadcasting Satellites (DBS)) en kabelnetwerken gestandaardiseerde MAC-systeem (MAC:"Multiplexed Analog Components") maakt een dergelijke evolutie mogelijk. Een z.g. HD-MAC signaal is hierbij enerzijds geschikt voor ontvangst met standaard MAC apparatuur en anderzijds voor ontvangst met een speciaal HD-MAC toestel dat een beeld kan weergeven van HDTV-kwaliteit. HD-MAC kan gezien worden als een Europees initiatief (Philips, Bosch, Thomson, e.a.) als tegenhanger van het Japanse MUSE-("Multiple Sub-Nyquist Encoding") system voor HDTV, dat niet de geschetste compatibiliteit bezit. Desalniettemin moet Japan toch als pionier worden gezien op het gebied van de HDTV, die zijn eerste proefuitzendingen al achter de rug heeft en in 1990 met rechtstreekse uitzendingen op Japan begint via DBS. Daar komt nog bij dat de mogelijkheid bestaat dat HDTV de consument reeds eerder bereikt via compact (video) disk en videocassette recorders, waarbij aansluiting bij een oude standaard (compatibiliteitseis) in mindere mate een rol speelt.

Zowel MUSE als HD-MAC zijn voorbeelden van analoge HDTV distributietechnieken via DBS en kabel, waarbij bandbreedtereductietechnieken worden toegepast om de bandbreedte te reduceren tot ongeveer 8 MHz. Let wel dat binnen de TV-ontvanger het signaal direct wordt gedigitaliseerd voor verdere bewerking, ruiscorrecties, digitale opslag, etc. Genoemde systemen zijn in staat om een enorme verbetering in beeldkwaliteit te verschaffen t.o.v. de huidige standaardresolutie tv zolang er maar een niet-gestoord signaal aanwezig is. Analoge transmissie heeft echter te lijden van (additive) ruis, vervorming en reflecties. Via kabel kan de transmissiekwaliteit goed zijn, maar veel bestaande kabelsystemen zijn niet ontworpen om te voldoen aan de toegenomen kwaliteitsstandaard en hebben ook te lijden van ruis, overspraak, etc. Vandaar dat er op verscheidene plaatsen in de wereld (Bellcore, Heinrich Hertz Institut) onderzoek wordt verricht aan digitale HDTV, waarbij het data transport loopt via een geheel op glasvezeltechnologie gebaseerd breedbandig digitaal netwerk, het z.g. BISDN, dat in staat moet zijn naast HDTV een groot aantal andere (video) diensten aan te bieden. De vele voordelen van digitale transmissie, zoals immuniteit tegen ruis, de mogelijkheid van vercijfering van de data, de relatieve eenvoud van de apparatuur en de onderhoudsvriendelijkheid zijn reeds genoegzaam bekend.

In de volgende paragrafen komen diverse aspecten van digitale HDTV aan de orde, zoals HDTV systeemeisen, inpassing binnen BISDN, compatibiliteit en hierarchische structuren en coderingstechnieken voor HDTV.

II HDTV Systeemeisen

Uit psychovisuele experimenten, uitgevoerd door de Japanse omroeporganisatie NHK, is gebleken dat grotere en scherpere beelden de kijkervaring aanzienlijk kunnen versterken — in een vergelijkbare mate als bij de vervanging van zwart/wit televisie door kleurentelevisie. Met name de vergroting van de aspectverhouding (verhouding tussen de beeldbreedte en de beeldhoogte) verhoogt het gevoel ruimtelijk in de scene aanwezig te zijn. Met bestaande TV-standaarden is het echter niet goed mogelijk om voldoende beeldkwaliteit te verkrijgen bij vergroting van de beeldafmetingen. Immers vlakflikker, lijnflikker en "line-crawl", kruiskleur en kruishelderheid, en zichtbaarheid van de lijnstructuur zullen als zeer hinderlijk ervaren worden. (Davidse, 1988). Een eerste stap ter verbetering van de beeldkwaliteit is de CCIR (Comite Consultatif International des Radiocommunications) aanbeveling 601 uit 1982, bedoeld als standaard voor de professionele digitale tv-studio, de z.g. 4:2:2 digitale studiostandaard. Deze standaard moet langzamerhand de huidige analoge kleuren TV standaarden PAL, SECAM en NTSC gaan vervangen. In de Amerikaanse literatuur wordt deze standaard wel aangeduid met "Extended Quality TV" (EQTV).

De belangrijkste parameters van deze standaard voor zowel 525/60 en 625/50 systemen, die nog eens worden samengevat in Tabel 1, zijn:

Table 1: Digitale 4:2:2 studiostandaard (CCIR 601).

	525 lijnen 60 fields/s	625 lijnen 50 fields/s
Aantal bemonsteringen per lijn (totaal):		
luminantie	858	864
chrominantie	429	432
Bemonsterfrequentie:		
luminantie	13.5 MHz	
chrominantie	6.75 MHz	
Aantal bemonsteringen per actieve digitale lijn:		
luminantie	720	
chrominantie	360	
Correspondentie tussen signaal- en kwantisatieniveau: (op een schaal van 0-255)		
luminantie	16 (zwart) tot 235 (wit)	
chrominantie	128 (geen kleur) ±112 (dwz. 16 tot 240 volledige verzadiging)	

- gescheiden codering van de videocomponenten: luminantie Y en kleurverschilsignalen (chrominanties) U en V.

- bemonsteringsfrequentie van 13.5 MHz voor de luminantie en 6.75 MHz voor beide chrominanties.

- uniforme kwantisering van 8 bits/bemonstering.

- transmissiesnelheid op basis van PCM: 8(13.5 + 6.75 + 6.75) = 216 Mbit/s.

Binnen deze nieuwe distributiestandaard worden zeer hoge eisen aan de beeldkwaliteit gesteld. Zichtbare vervorming door bandbreedte reducerende coderingstechnieken is niet toegestaan. Gaan we uit van 576 aktieve lijnen voor de luminantie Y met 720 beeldpunten per lijn en van 576 aktieve lijnen voor de chrominanties U en V met 360 beeldpunten per lijn, dan bedraagt de transmissiesnelheid van de feitelijke beeldinhoud bij 8 bits per beeldpunt en 25 beelden per seconde: (720 x 576 x 8 + 2 x 360 x 576 x 8) x 25 = 166 Mbit/s. Het verschil 216 - 166 = 50 Mbit/s is bedoeld voor synchronisatie-doeleinden.

De eisen die aan HDTV worden gesteld zijn ruwweg de volgende:

- Een vergroting van het scheidend (oplossend) vermogen (resolutie) in horizontale en vertikale richting met ongeveer een factor twee t.o.v. de digitale studiostandaard. Dit betekent een bitsnelheidsverhoging met een factor 6 tot 1.2 Gbit/s.

- Opheffing van de ongewenste onderlinge beinvloeding van kleur- en helderheidsinformatie (is reeds als eis verwoord in CCIR-601).

- Een vergroting van de aspectverhouding van 4:3 tot 16:9 (schermbreedte neemt met 33 % toe) om een beter aansluiting bij de filmwereld en de eigenschappen van het menselijk visueel systeem te verkrijgen.

- Stereofonische geluid met "hifi" kwaliteit.

In Tabel 2 wordt het CCIR-601 formaat vergeleken met het (nog niet gestandaardiseerde) HDTV-formaat. De keuze van 1125 lijnen bij het MUSE systeem is niet zo maar gekozen: met een decimatie van 9:5 krijgen we 625 lijnen van het PAL systeem en met een decimatie van 15:7 de 525 lijnen van het NTSC systeem.

De hoogwaardige beeldkwaliteit van HDTV (vergelijkbaar met 35 mm film) was in eerste instantie bedoeld voor omroep tv en videodistributie. Hoewel omroep nog steeds als belangrijke toepassing van HDTV wordt gezien, kan door de introduktie van een breedbandig digitaal netwerk (BISDN) een groot aantal andere toepassingen tegemoet gezien worden. Te denken valt hierbij aan een "Picture Archiving and Communication System" (PACS) gebaseerd op HDTV, waarbij medische specialisten hoogwaardige beelden kunnen opvragen en zichtbaar maken vanuit verafgelegen medische databanken. Door de superieure beeldkwaliteit zal dit systeem gemakkelijker ingang vinden binnen de medische wereld, dan de conventionele PACS systemen. Ook kunnen we denken aan videovergaderen en tele-educatie ondersteund met interactief op te vragen getypte tekst, hoogwaardig fotomateriaal, grafisch- en tekenmateriaal met veel detail.

III Het Breedbandig ISDN

Wereldwijd worden er pogingen ondernomen om belangstelling te wekken voor een nieuw op glasvezeltechnologie gebaseerd breedbandig multifunctioneel netwerk, bestemd voor de telecommunicatiediensten van morgen. Dit concept wordt breedbandig ISDN genoemd en wordt door PTT's en producenten van telecommunicatieapparatuur bestempeld als het netwerk van de toekomst, niet alleen vanwege de toegemeten brede band per individuele abonnee, maar vooral door de veelheid van diensten die kunnen worden aangeboden, varierend van geluid, data, video tot LAN (local area network) en MAN (metropolitan area network) en dit alles via een enkele 2-aderige glasvezel per abonnee. Van alle genoemde diensten zal waarschijnlijk het videotransport het meeste gebaat zijn bij dit breedbandige netwerk. Bij een hoogste bitsnelheid van 135 Mbit/s (BISDN H4) kan studio-kwaliteitvideo volgens CCIR-601 gemakkelijk worden gerealiseerd (EQTV). We zullen in de loop van de paragraaf aannemelijk maken, dat ook digitale HDTV die op hoogwaardige wijze is gecomprimeerd bij deze bitsnelheid kan worden verzonden.

De meeste BISDN zijn nog in een experimenteel stadium (Heinrich Hertz Institute, Bellcore). In Fig. 1 is de structuur geschetst van zo'n toekomstig BISDN in de vorm van een ster/ster architectuur. Verkeer vanuit de woonomgeving en kleine bedrijven wordt geconcentreerd in een z.g. Remote Electronics (RE) site en vervolgens samengevoegd met verkeer van grote bedrijven in een Central Office (CO). Door iedere abonnee een eigen glasvezel te geven, verkrijgen we een flexibele architectuur met voldoende privacy en security voor de abonnee. Het dynamisch bereik in bitsnelheid voor de verschillende diensten ligt in de orde van $0 - 10^8$ bit/s. We hebben te maken met een mengeling van

Table 2: Formaat CCIR-601 versus HDTV-formaat.

	CCIR 601	HDTV (MUSE)	HD-MAC-50
Aspect ratio	4:3	16:9	16:9
Aantal lijnen per frame	525/625	1125	1250
Aantal fields/s	59.94/50	60	50
Interliniering	2:1	2:1	2:1

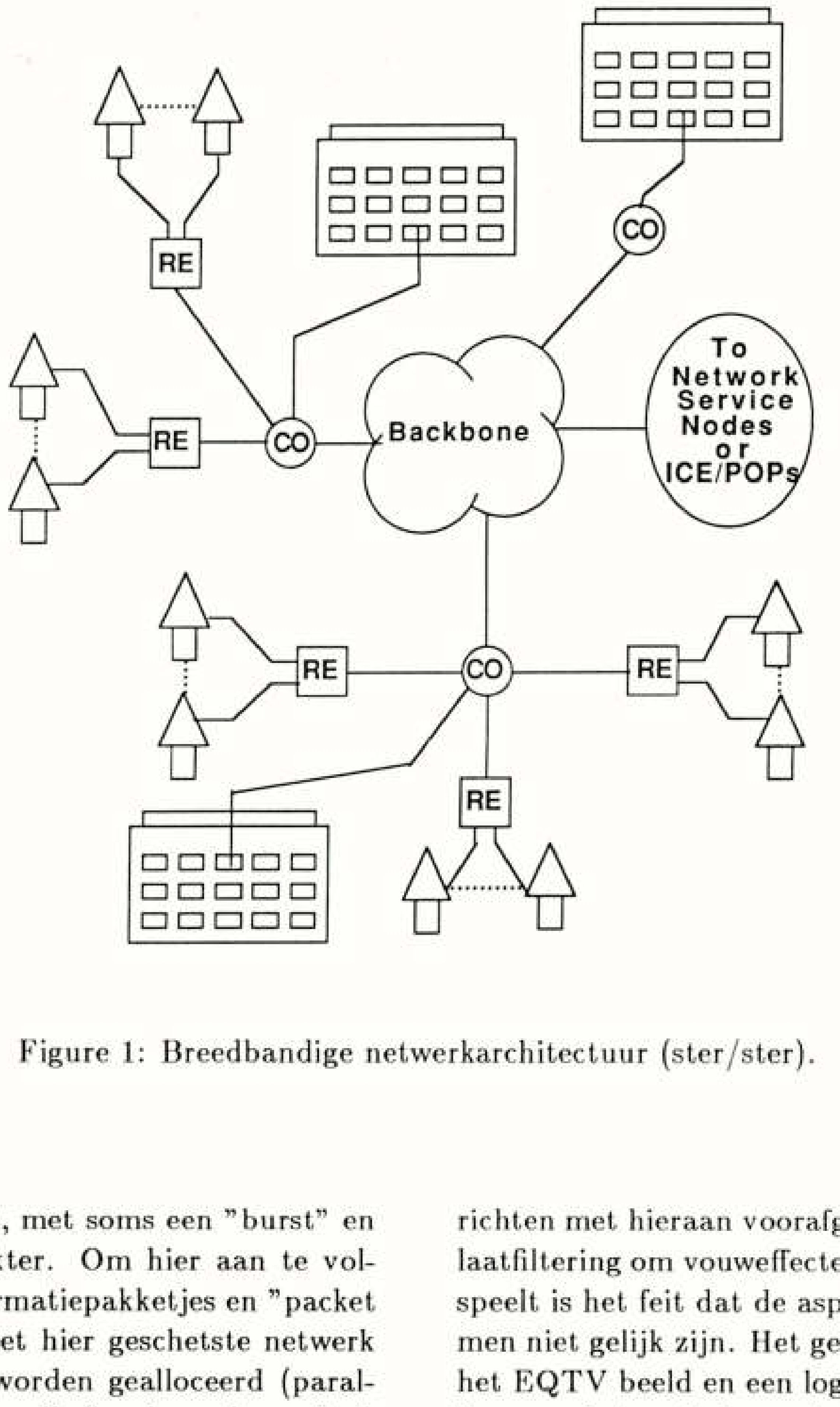


Figure 1: Breedbandige netwerkarchitectuur (ster/ster).

breedbandige en smalbandige diensten, met soms een "burst" en dan weer met een vrij continue karakter. Om hier aan te voldoen, wordt gebruik gemaakt van informatiepakketjes en "packet switching" technieken. Kan volgens het hier geschetste netwerk een veelheid aan diensten op afroep worden gealloceerd (parallel), bij HDTV wordt in een keer de volledige kanaalcapaciteit benut (135 Mbit/s). Vandaar dat we nog verder in de toekomst glasvezels verwachten met een kanaalcapaciteit van ongeveer 600 Mbit/s, waarmee meer-kanaals HDTV kan worden ontvangen.

IV Compatibiliteit en Hierarchische Structuren

In een omgeving waarbij veel beeldformaten een rol spelen, is het nuttig deze verschillende formaten zodanig te definieren dat conversie van de ene maat naar de andere eenvoudig mogelijk is. Binnen de BISDN omgeving hebben we zowel te maken met EQTV signalen als met HDTV signalen, zodat een eenvoudige conversie van het ene naar het andere formaat van groot belang is (compatibiliteit). Omdat het oplossend vermogen van een HDTV signaal ongeveer twee keer zo groot is in zowel horizontale als vertikale richting als dat van een EQTV signaal, kan zeer eenvoudig een hierarchische relatie tussen deze twee formaten gedefinieerd worden met EQTV als een deelverzameling van het HDTV signaal. Een manier om dit te bereiken is door het EQTV signaal af te leiden van het HDTV signaal door een 2:1 decimatie in beide richten met hieraan voorafgaand een twee-dimensionale laagdoorlaatfiltering om vouweffecten te vermijden. Een probleem dat hier speelt is het feit dat de aspect verhoudingen voor de beide systemen niet gelijk zijn. Het gedecimeerde HDTV beeld is breder dan het EQTV beeld en een logische keus zou zijn om de twee zijflanken van het gedecimeerde HDTV beeld er af te knippen zodat het overblijvende signaal past binnen de 4:3 aspect verhouding. Een andere mogelijkheid is het verschaffen van een stuursignaal waarmee het gewenste gebied kan worden geselecteerd. Dit wordt vaak aangeduid met "pan and scan".

Worden zowel HDTV en EQTV diensten aangeboden op het BISDN, op basis van de geschetste hierarchie, dan kan een abonnee met slechts een EQTV decoder het EQTV signaal ontvangen door de noodzakelijke deelverzameling uit de 135 Mbit/s bitreeks te extraheren. De abonnee heeft dus de keuze om een programma in EQTV formaat te bekijken bij een lagere bitsnelheid, bijv. 45 Mbit/s, terwijl anderen het op een breed scherm bekijken bij volle bandbreedte en resolutie.

De vereiste compatibiliteit tussen EQTV en HDTV kan op verschillende manieren worden bereikt. In het hier nu volgende zullen we kort ingaan op drie z.g. hierarchische coderingsschema's.

Schema I: Een eenvoudige methode om tot een hierarchische codering te komen is door gebruik te maken van de pyramide structuur, die tevens geschikt is voor hardware implementatie. In Fig. 2 wordt een 2-niveaus pyramide structuur getoond om een compatibiliteit te realiseren tussen HDTV en EQTV. In het

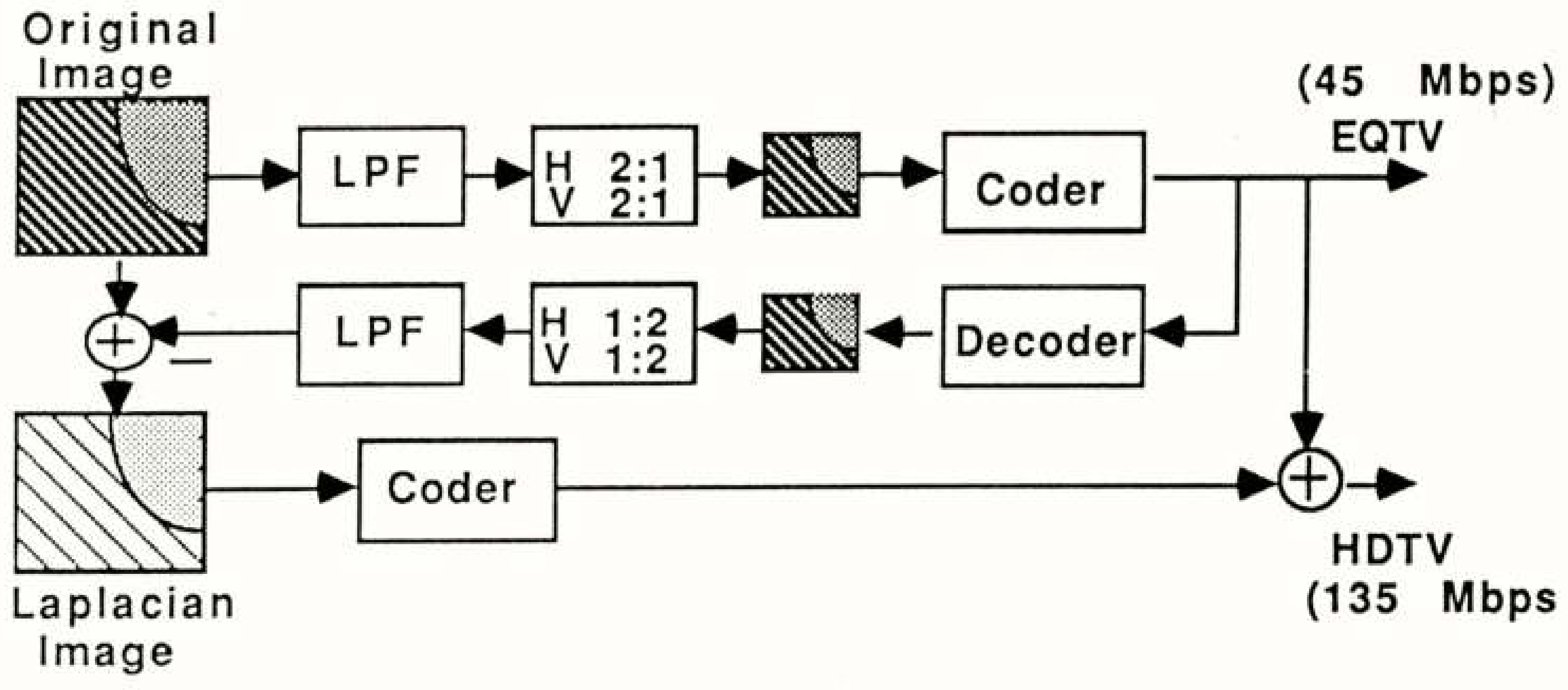


Figure 2: Hierarchische codering op basis van pyramide structuur.

eerste pad wordt het inkomende HDTV signaal laagdoorlatend gefilterd en gedecimeerd in horizontale en vertikale richting (2:1) tot een kwart van zijn oorspronkelijke afmeting. Het gedecimeerde signaal dient dan als basis voor het EQTV signaal en wordt gecodeerd met 45 Mbit/s. Het gedecodeerde EQTV signaal wordt na interpolatie teruggebracht op de oorspronkelijke beeldafmeting en wordt vervolgens afgetrokken van het oorspronkelijk HDTV signaal. Het resulterende verschilsignaal in het tweede pad levert tesamen met het EQTV signaal uit het eerste pad het HDTV signaal van 135 Mbit/s. We zien dat de informatie die tijdens het coderen van het EQTV signaal verloren ging, weer gereconstrueerd kan worden uit het verschilsignaal.

Hoewel dit schema relatief eenvoudig te implementeren is en vrij robuust is, kan niet gezegd worden dat de datarepresentatie erg efficient is voor coderingsdoeleinden. Immers de hoeveelheid te coderen data is 25% (gedecimeerd beeld) + 100% (verschilbeeld) = 125%. Het is dan ook efficienter om in plaats van een pyramide structuur voor hierarchische codering uit te gaan van een subband coderingsstructuur, waarin voorafgaande aan de feitelijke codering geen datavermeerdering plaats vindt.

Schema II: In Fig. 3 wordt een hierarchisch coderingsschema getoond met subband splitsing. Door gebruik te maken van 2-D scheidbare Quadrature Mirror Filters (QMF's) kunnen we het beeld opsplitsen in 4 subbanden. Na decimatie is het laagdoorlaat gefilterde en gedecimeerde beeld een lage resolutie replica van het HDTV beeld. Deze basisband kan dan worden gebruikt om het gecodeerde EQTV signaal af te leiden met een bitsnelheid van 45 Mbit/s. Tengevolge van deze coderingsoperatie gaat er nogal wat hoogfrequent informatie verloren uit de basisband die bij reconstructie van het HDTV signaal uit de 4 subbanden (inclusief de basisband) node gemist wordt. Merk op dat in Schema I deze verloren gegane informatie weer gereconstrueerd kan worden uit het verschilsignaal. Om dit probleem het hoofd te bieden, wordt het coderen van de basisband opgedeeld in twee paden (progressieve codering): het uitgangssignaal van pad 1 verschaft het EQTV signaal, terwijl het uitgangssignaal van pad 2 dat additionele informatie bevat, wordt gebruikt om te worden gecombineerd met de andere banden om zodoende het HDTV signaal te reconstrueren. Op deze wijze kan een uitstekende HDTV beeldkwaliteit worden bereikt (Biemond et al., 1990).

Schema III: Een andere techniek die zich goed leent om te worden toegepast binnen een hierarchische coderingsomgeving is transformatiecodering m.b.v. de DCT. Zowel het filteren als de decimatie kan worden uitgevoerd in het transformatiedomein. Er kan worden bewezen (Schipper, 1988) dat de DCT voldoet aan een gemodificeerde vermenigvuldiging-convolutie eigenschap. Dit betekent dat de laagdoorlaatfiltering in het transformatiedomein kan worden uitgevoerd door elke transformatiecoefficient te vermenigvuldigen met een weegfactor overeenkomstig de gewenste "spectrale" vorm en dat de decimatie kan worden uitgevoerd door de inverse transformatie uit te voeren op een gereduceerde verzameling coefficienten beneden de afsnijfrequentie resulterend in een gedecimeerd beeld. Omdat de rekeninspanning van een discrete transformatie toeneemt met de afmeting van de transformatie evenals de complexiteit van de vereiste raster-blok conversie, wordt een beeld meestal opgedeeld in blokken van 8 x 8, 16 x 16 of 32 x 32 beeldelementen. Dit betekent dat de filtering en de decimatie bloksgewijs in het transformatiedomein worden uitgevoerd.

Fig. 4 toont het hierarchische coderingsschema met gebruikmaking van de DCT voor de EQTV signaal-extractie. De structuur van het schema is eenvoudiger dan de voorgaande systemen. Daar staat echter een vrij complex coderingsmechanisme tegenover. Immers ook hier geldt dat het EQTV signaal ter aanpassing aan het 45 Mbit/s kanaal nogal wat hoogfrequent informatie heeft verloren die voor de reconstructie van het HDTV signaal onontbeerlijk is. Vandaar dat we ook hier te maken hebben met twee paden. Het eerste bevat het afgevlakte EQTV signaal op 45 Mbit/s en het tweede pad bevat de aanvullende verfijningsinformatie. Een dergelijke techniek waarbij een beeld aanvankelijk wordt uitgezonden met een lagere bitsnelheid en verfijnd wordt door additionele informatie noemen we progressieve codering. De kwaliteitstoename kan worden bereikt door additionele informatie via een verfijning van de reeds gekwantiseerde transformatiecoefficienten en door het toevoegen van extra transformatiecoefficienten door bijvoorbeeld de drempel bij het drempelcoderen te verlagen.

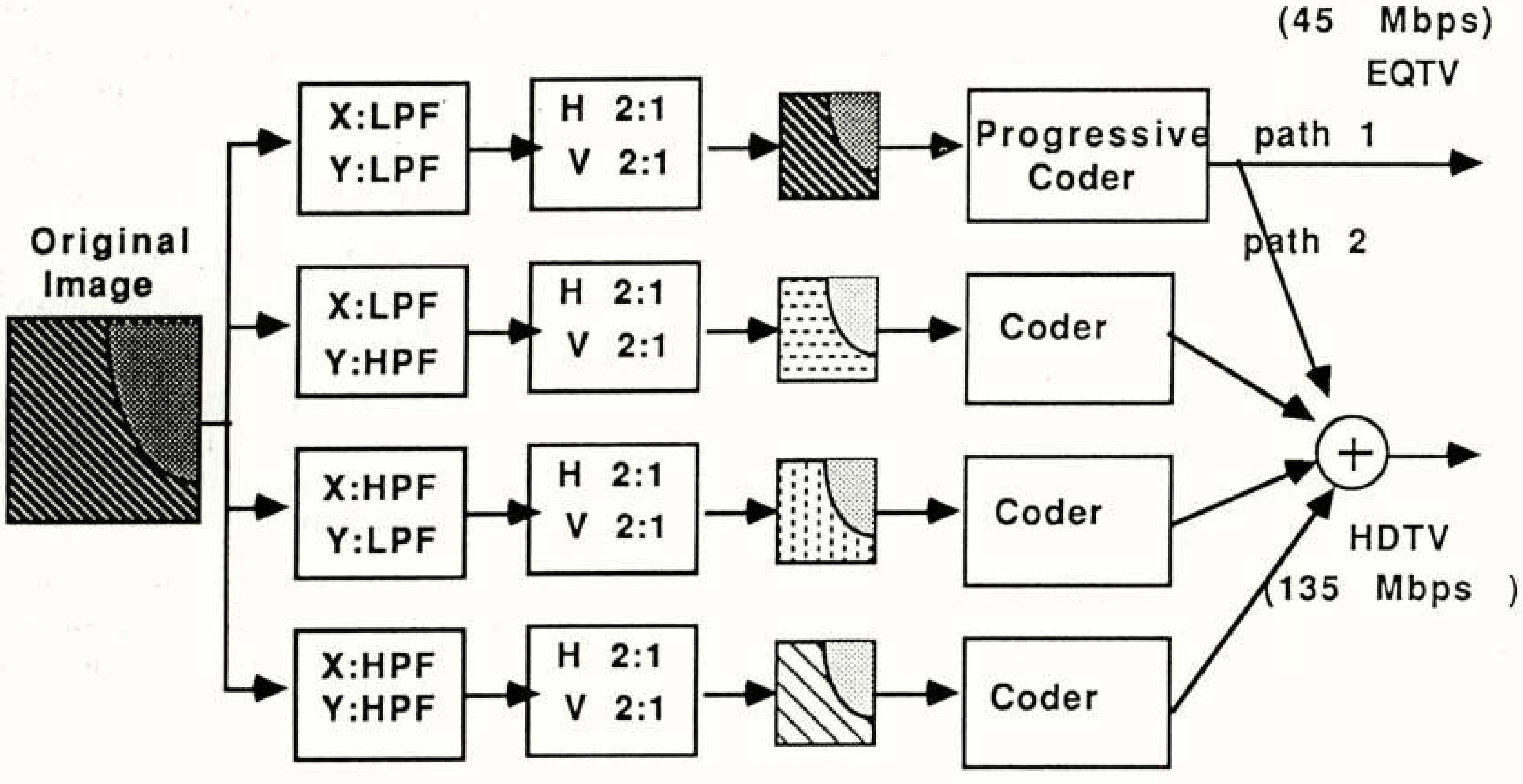


Figure 3: Hierarchische codering op basis van subband splitsing.

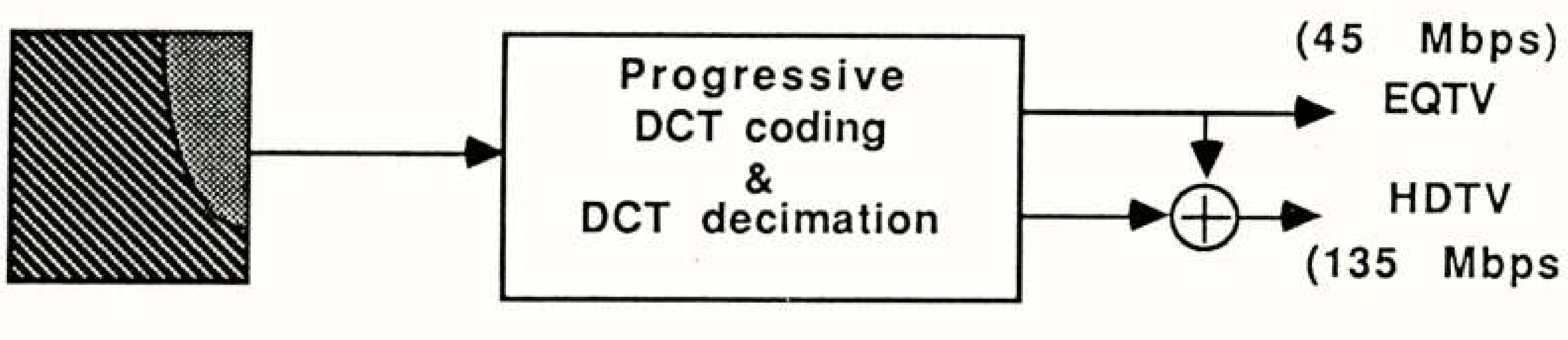


Figure 4: Hierarchisch coderingssysteem gebaseerd op de DCT, progressieve transmissie en decimatie.

V Coderingstechnieken voor HDTV

De bemonsteringsfrequentie voor de verschillende voorgestelde HDTV systemen ligt tussen de 48 en 75 MHz. Dit betekent voor HDTV codering op 135 Mbit/s een gemiddelde bitsnelheid van 2.8 tot 1.8 bit per kleurenbeeldpunt. Theoretisch gesproken kunnen alle technieken die gebruikt worden voor het coderen van digitale video bij lage bitsnelheden ook gebruikt worden voor het coderen van HDTV. Echter, niet ieder algoritme is even geschikt voor digitale HDTV vanwege de zeer korte verwerkingstijd die vereist is t.g.v. de zeer hoge bemonsteringssnelheid. Deze beperking zal minder zwaar gaan wegen naarmate er meer vooruitgang wordt geboekt op het gebied van snelle hardware.

Een effektieve bitreductietechniek is bewegingsadaptieve spatieel/temporele sub-Nyquist bemonstering. Deze techniek vormt de basis van de analoge MUSE en HD-MAC systemen. In deze benadering wordt datacompressie bereikt door spatiele onderbemonstering in bewegende gebieden en temporele onderbemonstering in stationaire en langzaam bewegende gebieden. Dit idee is gebaseerd op de eigenschap van het menselijk visueel systeem dat bewegende objecten met een geringere spatiele resolutie worden waargenomen dan stilstaande objecten. Zodoende zal spatiele onderbemonstering in veel gevallen geen noemenswaardige degradaties geven. In scenes met geringe beweging, waar het oog de bewegende objecten kan volgen is er echter wel degelijk sprake van zichtbare vervorming.

Naast de sub-Nyquist bemonsteringstechniek, wordt ook DPCM als een geschikt algoritme gezien voor HDTV toepassingen. Het hierarchische coderingsschema (schema I) uit Fig. 2 wordt hiertoe gebruikt omdat het slechts twee data paden bezit, het minimale vereiste voor compatibele codering. Deze structuur minimaliseert de hardware complexiteit, die evenredig is met het aantal paden.

Ten tweede zorgt de terugkoppeling ervoor dat informatie die in het eerste pad verloren is gegaan, danwel dat coderingsfouten worden gecorrigeerd in het tweede pad. We hebben reeds opgemerkt dat de gebezigde pyramide structuur aanleiding geeft tot een datavermeerdering met 25%. In Fig. 5 is schema I nader uitgewerkt met een intra-frame DPCM schema in pad 1 en entropie codering in pad 2. Doordat de bemonsterfrequentie in het eerste pad met een factor 4 is gereduceerd, is de DPCM verwerkingstijd ongeveer gelijk aan die van een conventionele EQTV coder, ongeveer 75 ns. De 45 Mbit/s van het EQTV signaal wordt gemultiplexed met het 90 Mbit/s verschilsignaal t.b.v. de transmissie. Aan de ontvangstzijde wordt de eerste 45 Mbit/s bitstroom gebruikt om het EQTV signaal te reconstrueren, terwijl de combinatie van 45 Mbit/s en 90 Mbit/s data stromen worden gebruikt om het HDTV signaal te reconstrueren.

Schema II (subband codering) en Schema III (DCT codering) zullen hier niet nader worden uitgewerkt. Ze zijn momenteel het onderwerp van wereldwijd onderzoek omdat ze een hoge compressie paren aan een relatief goede beeldkwaliteit. Daar staat tegenover dat implementatie van deze technieken in real-time hardware nog wel even op zich zal laten wachten.

References

[Annegarn] M.J.J.C. Annegarn et al., "HD-MAC: Een stap vooruit in de evolutie van de televisietechniek", Philips Technisch Tijdschrift, Jaargang 43, no, 8, 1986.

[Biemond] J. Biemond, F. Bosveld and R.L. Lagendijk, "Hierarchical Coding of HDTV in BISDN", Proc. 1990 IEEE Int. Conf. on Acoustics, Speech and Signal Processing, Albuquerque NM.

[Bosveld] F. Bosveld, "Hierarchische Codering van HDTV in BISDN", Afstudeerverslag TU Delft, Fac. Elektrotechniek, Vakgroep Informatietheorie, Nov. 1989.

[Davidse] J. Davidse, Elektronische Beeldtechniek, Collegediktaat TU Delft, Fac. Elektrotechniek, 1988.

[Schipper] R. Schipper, Resolutieconversie van Digitale Beelden, Afstudeerverslag TU Delft, Fac. Elektrotechniek, Vakgroep Informatietheorie, 1988.

[Bellisio] J.A. Bellisio and K-H. Tzou, "HDTV and the Emerging Broadband ISDN Network", Proc. SPIE Int. Conf. on Visual Communications and Image Processing 1988, Boston, MA, pp. 772–786.

[Ninomiya] Y. Ninomiya et al. "An HDTV Broadcasting System Utilizing a Bandwidth Compression Technique — MUSE", IEEE Trans. on Broadcasting, vol. BC-33, no. 4, Dec. 1987, pp. 130–143.

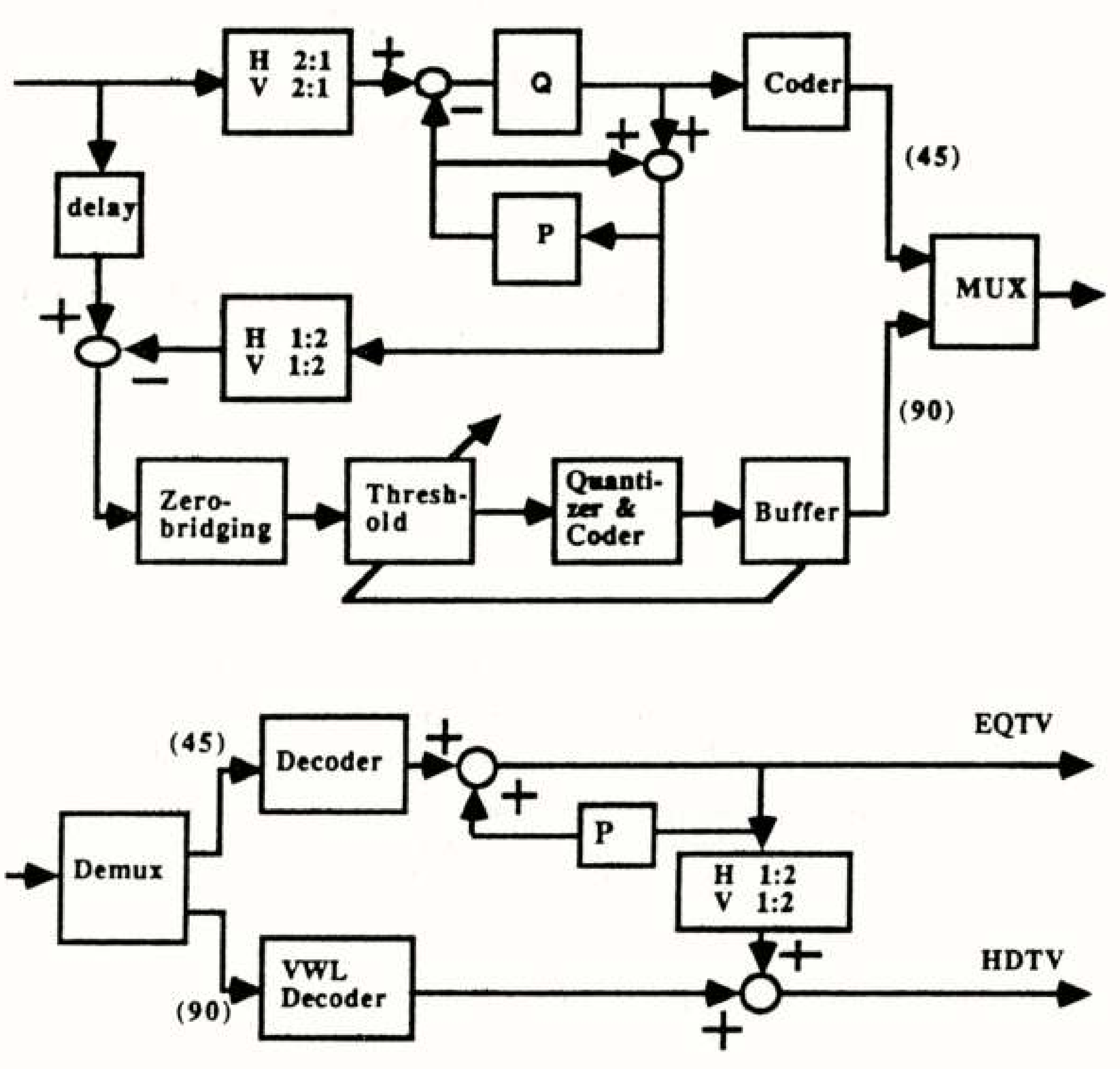
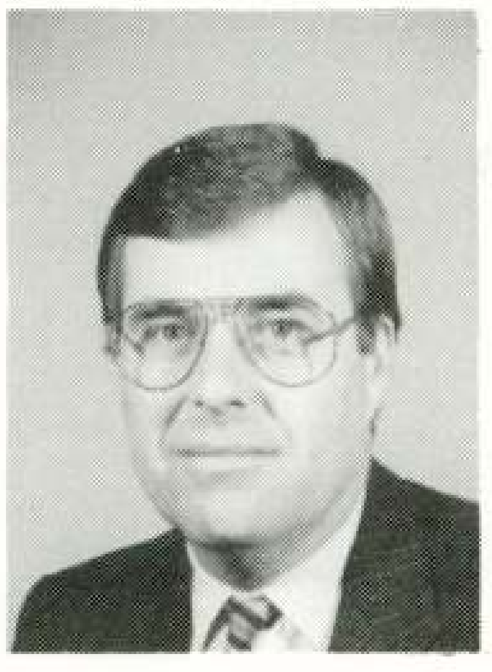


Figure 5: Een hierarchische structuur voor compatibele HDTV codering (a) encoder; (b) decoder.

Voordracht gehouden tijdens de 374e werkvergadering.

NEDERLANDS ELEKTRONICA- EN RADIOGENOOTSCHAP
(374ste werkvergadering)
IEEE BENELUX SECTIE
AFDELING TELECOMMUNICATIE KIVI
AES



Prof.Dr.Ir. J. Biemond



Dr.Ir. M. Breeuwer

UITNODIGING

Voor de lezingendag op donderdag 7 december 1989 in de collegezaal WB parterre van het Philips Natuurkundig Laboratorium, Prof. Holstlaan, Eindhoven.

Thema: Digitaal Video

Programma

10.30 - 11.00 uur	Ontvangst, koffie en welkomswoord door Drs. M. Carasso, (Directeur Natuurkundig Laboratorium Philips)
11.00 - 11.45 uur	**Prof.Dr.Ir. J. Biemond,** (TU-Delft "Digitaal Video: Een Overzicht").
11.45 - 12.30 uur	**Dr.Ir. M. Breeuwer,** (Philips Natuurkundig Laboratorium, Eindhoven): "Broncodering van TV-en HDTV-signalen".
12.30 uur	Lunch, aangeboden door Philips.
14.00 - 14.45 uur	**Ir. R. Plompen,** (PTT, Neher Lab.): "Broncodering voor beeldtelefonie".
14.45 - 15.15 uur	Thee.
15.15 - 16.00 uur	**Dr.Ir. A. van Roermund,** (Philips Natuurkundig Laboratorium, Eindhoven): "Real-Time Video: van idee tot IC".

Aanmelding voor de lezingen dient te geschieden vóór 25 november 1989 door middel van aangehechte kaart, gefrankeerd met een postzegel van 55 cent. Het Natuurkundig Laboratorium is bereikbaar met bus 177 die vertrek vanaf de oostzijde van de noorduitgang van het station Eindhoven.

Namens het IEEE bestuur,

Prof.Dr.Ir. J.B.H. Peek,
Tel. 040-742451

Instrumentatie voor radioastronomisch spectraallijn onderzoek

Dr.Ir. A. Bos

Stichting ASTRON, Dwingeloo

Instrumentation for radioastronomical spectral line observations.
This contribution gives a description of the spectrometers used for spectral line research in radioastronomy. The digital correlation spectrometers currently in use at the Westerbork Synthesis Radio Telescope are briefly described followed by an example of current developments within the Netherlands Foundation for Research in Astronomy: the construction of a spectrometer for submillimeter spectroscopy with an effective bandwidth of 1 GHz and a resolution of 1.25 MHz for which NFRA developed a custom integrated digital correlator circuit.

1. Inleiding

Spectraallijn onderzoek betreft in het algemeen het meten van de intensiteit, spectrum en richting van de straling uitgezonden door de waargenomen objecten. Het spectrum verschaft enerzijds informatie over de snelheidsverdeling binnen, anderzijds informatie over de chemische samenstelling van deze objecten.

De instrumentatie voor het spectraallijn onderzoek, voorzover dat onder de verantwoordelijkheid van de Stichting ASTRON valt, betreft drie instrumenten:

- De 25 m Dwingeloo telescoop waarmee voornamelijk onderzoek wordt gedaan aan de lijnstraling van atomaire waterstof en de OH radicaal op respectievelijk 21 en 18 cm golflengte.
- De Westerbork Synthese Radio Telescoop (WSRT), bestaat uit een 14-tal parabolische antennes met een diameter van 25 m die uitgerust zijn met een ontvanger systeem voor 92, 50, 21 en 6 cm golflengte.
  De WSRT wordt gebruikt voor het waarnemen met hoog ruimtelijk oplossend vermogen van breedband continuumstraling en relatief smalband lijnstraling op de genoemde golflengtes.
- De James Clerk Maxwell Telescoop (JCMT) is een 15 m antenne voor het submillimeter golflengtegebied die op Hawaii is opgesteld op een hoogte van 4200 m. Deze telescoop wordt gebruikt voor het spectraallijn onderzoek aan complexe moleculen (HCN, CO, etc.) op (sub)mm golflengtegebied.

De ontvangersystemen voor deze telescopen bestaan uit een heterodyne ontvanger waarvan de detector bestaat uit een spectrometer. De spectrometer voor een synthese telescoop verschilt van die voor een enkele telescoop in de wijze waarop naast de spectrale informatie de ruimtelijke informatie gemeten wordt. De 14 antennes van de WSRT zijn opgesteld op een Oost-West lijn met een lengte van 3 km. Gebruik makend van de rotatie van de aarde wordt de correlatie gemeten tussen nominaal 40 antennecombinaties. Zo wordt in 12 uur, na een 3-dimensionale Fouriertransform, een "antenne" gesimuleerd met een effectieve diameter van ongeveer 3 km.

2. Spectrometer typen

Voor de spectrometers gelden een aantal eisen die samenhangen met de zeer lage intensiteit en de diversiteit van de waargenomen lijnprofielen.
De lage intensiteit vergt veelal lange integratietijden.
Voor een spectrometer geldt dat de r.m.s. ruis per punt op het waargenomen lijnprofiel gegeven is door:

$$\Delta T_{rms} \div T_{sys} \cdot \frac{d}{\sqrt{\Delta B T}} \qquad (2.1)$$

Waarin T_{sys} = de systeem ruis temperatuur in K.
ΔB = de effectieve bandbreedte per gemeten punt (Hz).
T = de integratie tijd (sec).
d = een degradatie factor die afhangt van het type spectrometer en de waarneemprocedure.

Om bijvoorbeeld de vorm van een spectraallijn met een piekintensiteit van 1 K en bij een ontvangerbandbreedte van 5 MHz met een resolutie van 1000 punten voldoende nauwkeurig te bepalen is een r.m.s. ruis niveau vereist van 0.01 K. Bij een systeem ruistemperatuur van 30 K is dan een integratietijd nodig van ongeveer 1800 sec. De overdrachtsfunctie van het systeem dient gedurende deze tijd stabiel te blijven op het niveau <0.01 K (1:3000). Voor een enkele telescoop betreft een dergelijke eis alleen de amplitude. Voor een synthese telescoop is veelal een amplitudestabiliteit van 0.1% en een fasestabiliteit van 1 graad vereist over een periode van 12 uur [1]. In de praktijk wordt de effectieve stabiliteit, afhankelijk van de tijdschaal van de

variaties, verbeterd ten koste van waarneemtijd door het voldoende vaak meten van een referentiespectrum voor de calibratie van de doorlaatband.

De breedte van de waargenomen lijnprofielen hangt af van het stralings-mechanisme, de snelheidsverdeling binnen het waargenomen object en de golflengte. Zo varieert de bandbreedte nodig voor het waarnemen van een lijnprofiel op cm golflengte van 156 KHz voor een recombinatielijn op 92 cm tot 10 MHz voor galactische waterstof op 21 cm. Voor het (sub)mm gebied varieren de benodigde bandbreedtes tussen 100 MHz en 2 GHz.

Voor het voldoende nauwkeurig bepalen van de vorm en de plaats van het profiel (de snelheidsverdeling van de waargenomen straling) is een spectraal oplossend vermogen gewenst van 0.1 - 1 procent.

De technische eisen voor een spectrometer betreffen samengevat een zo goed mogelijke gevoeligheid, stabiliteit en flexibiliteit binnen een gegeven budget.

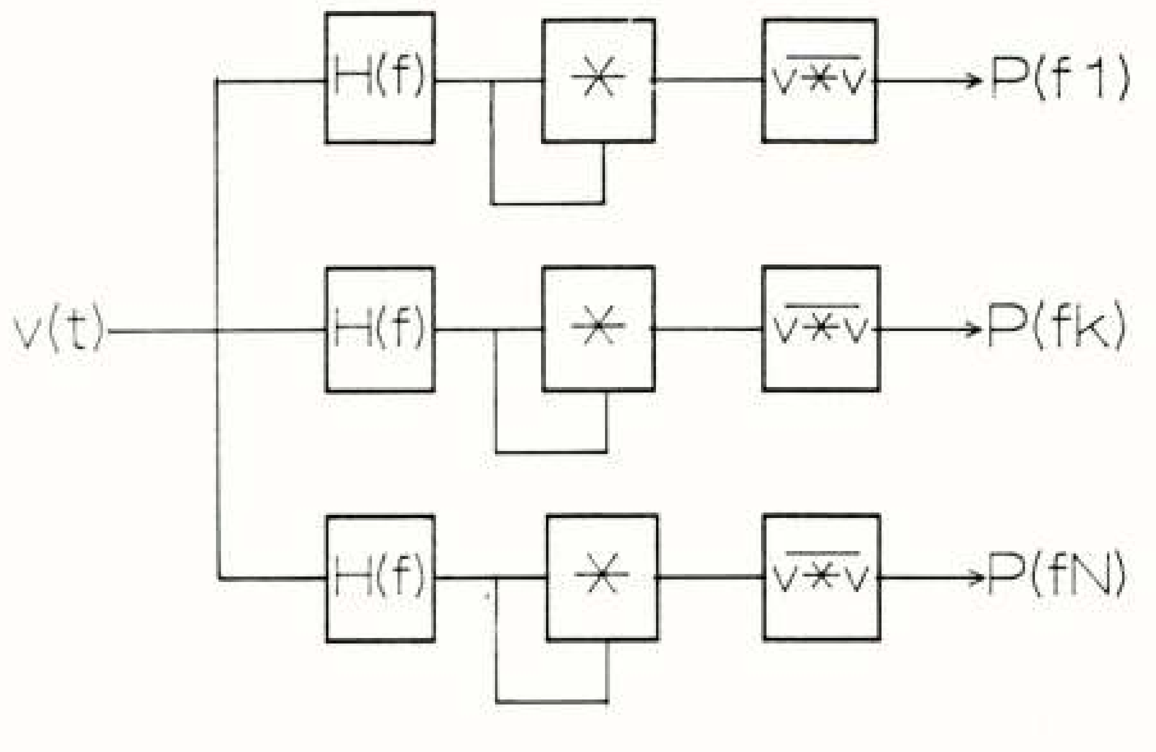


Fig. 2.1 Filterbank spectrometer

Binnen de radioastronomie zijn 5 typen spectrometers in gebruik:

a) de analoge filterbank spectrometer (zie Fig. 2.1.; [2]): de vorm van het (vermogens) spectrum wordt met behulp van, in het algemeen LC, filters gevolgd door een kwadratische detector en een integrator bepaald. Deze spectrometers worden voornamelijk toegepast in systemen met één enkele telescoop. Ze zijn relatief duur, hebben i.h.a. een vaste bandbreedte en het is moeilijk om met name de detectoren en de integratoren voldoende stabiel te maken en om het niveau van ongewenste mengproducten voldoende laag te houden.

b) de digitale auto- of kruiscorrelatie spectrometer (zie Fig. 2.2.; [3]): het spectrum wordt verkregen na de Fouriertransform van een op digitale wijze berekende correlatie functie. Het voordeel van deze correlatoren ligt in hun relatieve eenvoud en inherente stabiliteit. Dit is des te meer het geval omdat vanwege het Gaussisch karakter van de gemeten signalen in principe volstaan kan worden met 1-bit correlatie [4]. Dit introduceert echter een degradatie van de gevoeligheid met een factor 1,57 die gereduceerd kan worden tot een factor 1,14 door het gebruik van 2-bit correlatie [5].

Een digitale correlator (Fig. 2.2) bestaat uit een N-traps schuifregister waarvan de uitgangen worden toegevoerd aan de één van de twee ingangen van N vermenigvuldigers. De andere ingang is, in het geval van een autocorrelator, het ingangs-signaal van de eerste trap, voor een kruiscorrelator een, over ½N trappen vertraagd, signaal afkomstig van een tweede antenne. De N producten worden geintegreerd in N optellers. Het vermogens spectrum wordt verkregen na Fouriertransformatie van de gemeten correlatiefunctie.

De bandbreedte kan gemakkelijk verandert worden door wijziging van de klokfrekwentie.

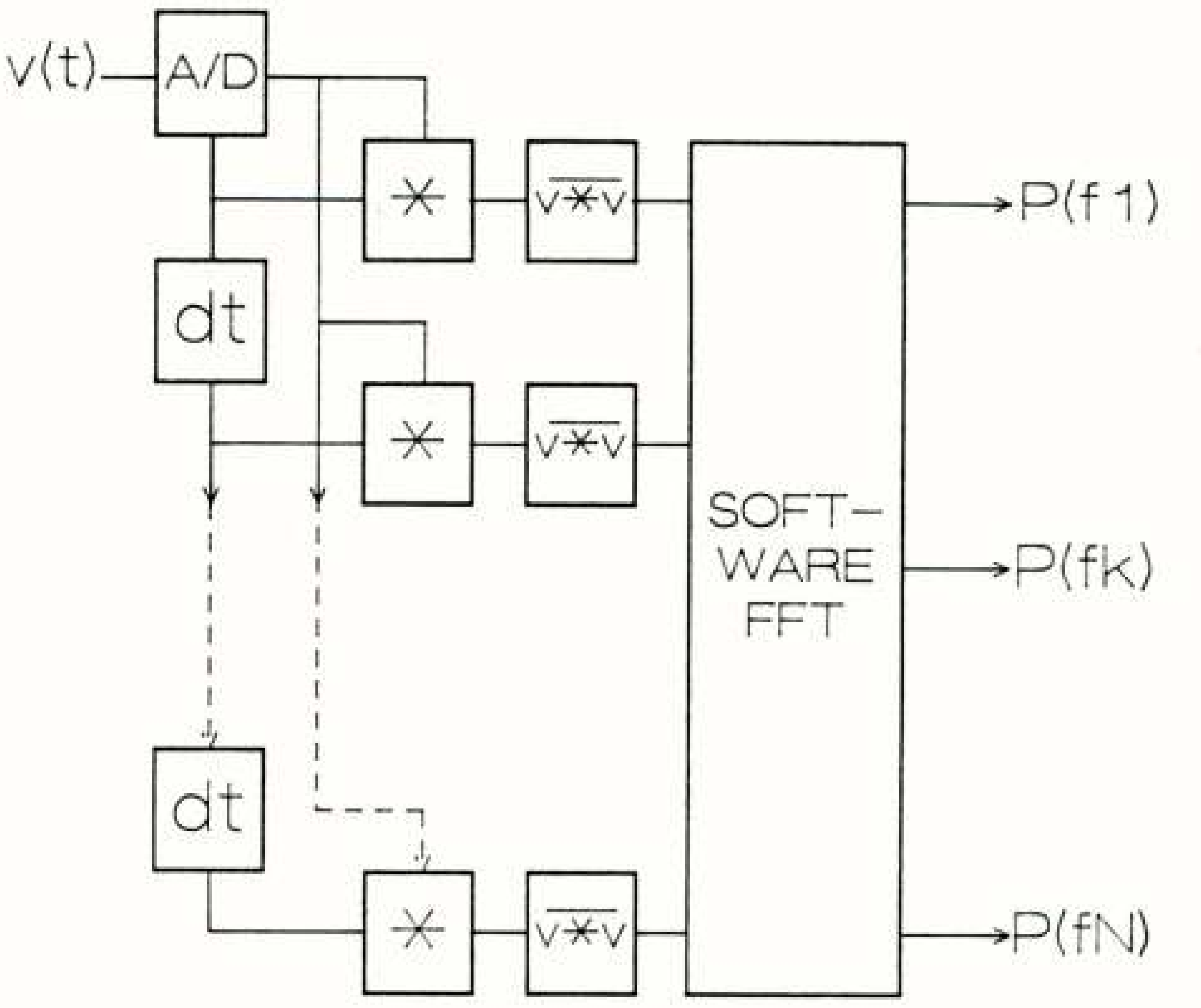


Fig. 2.2 Autocorrelatie spectrometer

c) de Fouriertransform spectrometer (zie Fig. 2.3.; [6]): dit is in feite een digitale filterbank spectrometer. In deze filterbank wordt een N-punts Fouriertransformatie uitgevoerd op segmenten van het ingangssignaal waarna, voor een enkele telescoop, het resultaat (digitaal) gekwadrateerd en geïntegreerd wordt. Voor een synthese telescoop is iedere antenne uitgerust met een dergelijke filterbank die dan een reëel naar complexe Fouriertransformatie uitvoert. Per antenne combinatie worden N complexe produkten gevormd die vervolgens geïntegreerd worden. Dit type spectrometer is inherent stabiel. De kosten zijn afhankelijk van het aantal bits dat in de FFT meegenomen wordt. De bandbreedte kan eenvoudig verandert worden.

d) de akoestoöptische spectrometer (AOS) (zie Fig. 2.4.; [7]): hierbij wordt gebruik gemaakt van een Bragg cel waarin het ingangssignaal een akoestisch

golfpatroon opwekt. Dit golfpatroon veroorzaakt frekwentie afhankelijke diffractie in een evenwijdige coherente monochromatische lichtbundel. Het vermogens spectrum wordt geproduceerd door een lineaire fotodetector, waarop het diffractie patroon wordt afgebeeld door een lens. De lens vervult hierbij de functie van de Fouriertransform; de fotodetector die van kwadratische detector en integrator.
Het voordeel van een AOS is dat de kosten relatief laag zijn. Een nadeel is de vaste bandbreedte en dat deze systemen in de praktijk niet bruikbaar zijn voor interferometrie. De bandbreedte van de beschikbare Bragg cellen ligt tussen 100 en 1000 MHz hetgeen deze systemen aantrekkelijk maakt voor spectroscopie op mm golflengtes. Een nadeel is echter de beperkte stabiliteit, het niet-lineair karakter van de frequentie as en de inflexibiliteit

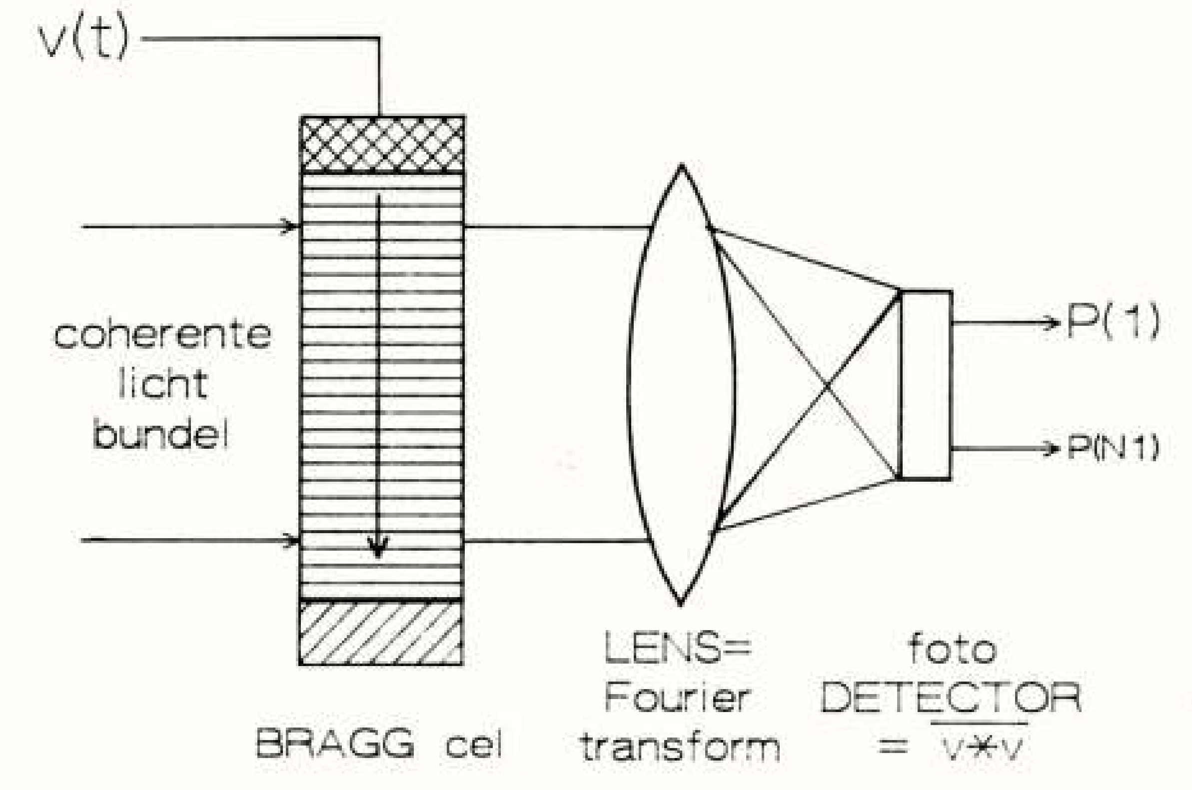


Fig. 2.4 Akoestooptische spectrometer

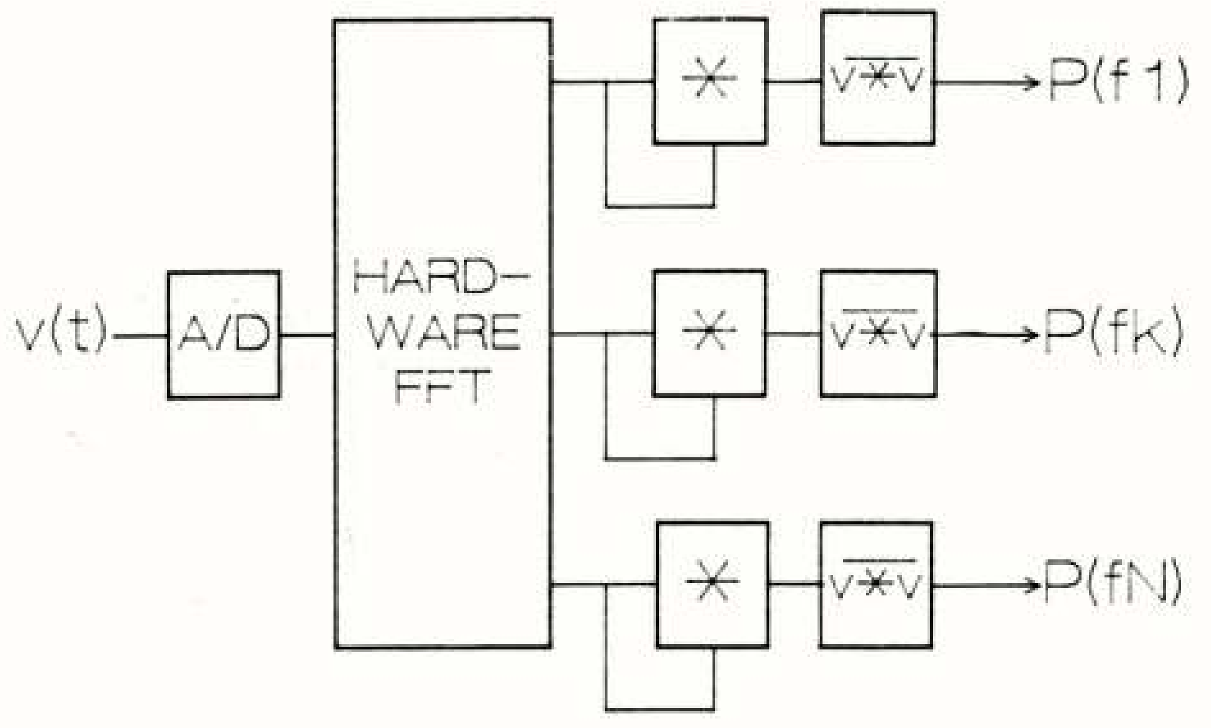


Fig. 2.3 Fouriertransform spectrometer

in de keuze van de bandbreedte. Het aantal frequentie punten is ongeveer 1000.

e) <u>een hybride oplossing</u> (zie Fig. 2.5.). Dit is een analoge filterbank gevolgd door een digitale correlator.
De capaciteit C van een autocorrelatie spectrometer die een N-punts spectrum over een totale bandbreedte B produceert is gegeven door:

$$C + 2\alpha NB$$

waarin α een factor is die afhangt van het aantal bits dat gecorreleerd wordt.
Wanneer de band B wordt opgedeeld in M subbanden met bandbreedte B/M is, voor eenzelfde spectrale resolutie van N punten over de totale band B, de benodigde correlator capaciteit gegeven door:

$$C + 2\alpha \frac{N}{M} \cdot \frac{B}{M} \cdot M = 2\alpha \frac{NB}{M} \qquad (2.2)$$

De benodigde correlator capaciteit is dus omgekeerd evenredig met het aantal subbanden. De kosten van de correlator zijn ruwweg evenredig met de correlator capaciteit. Voor het totale systeem nemen echter de kosten van de filterbank en de A/D omzetters ongeveer evenredig toe met M. Het optimum hangt allereerst sterk af van de grootte van B. Vervolgens spelen de eisen die gesteld moeten worden aan de stabiliteit en de onderdrukking van ongewenste mengproducten een grote rol. Het analoge deel dient om deze redenen zo beperkt mogelijk gehouden.

De keuze tussen de verschillende typen is afhankelijk van de toepassing.
In de praktijk worden alle genoemde typen in diverse vormen toegepast.
Voor de kleinere bandbreedtes (<40 MHz) worden veelal correlatie spectrometers gebruikt. Voor de grotere

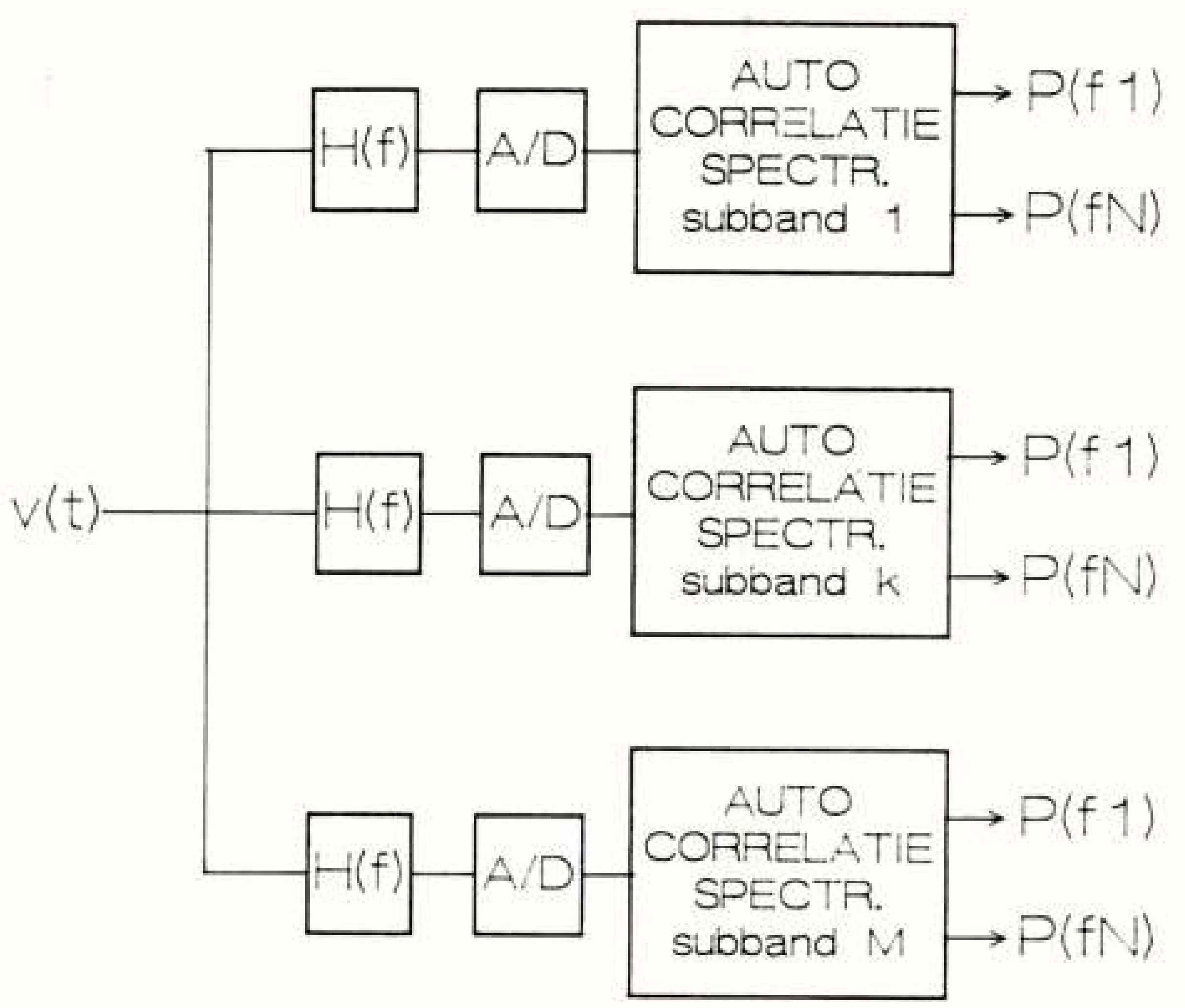


Fig. 2.5 Hybride spectrometer

bandbreedtes ligt het accent op akoestooptische en hybride spectrometers. Fouriertransform spectrometers zijn in de praktijk alleen goedkoper dan 1- of 2-bit correlatiespectrometers voor zeer grote systemen.

3. Correlatie spectrometers voor de WSRT

Voor de WSRT hebben we op dit ogenblik de beschikking over een tweetal digitale kruiscorrelatie spectrometers [8, 9, 10].

Het eerste systeem (zie Fig. 3.1.) wordt gebruikt voor het normale spectraallijn onderzoek. Het is een 2560 kanaals digitale kruiscorrelator (zie 2b) die op een, met een signaal bandbreedte B van 10 MHz corresponderende, kloksnelheid van 20 MHz werkt. Voor bandbreedtes kleiner dan 10 MHz heeft de correlator een overcapaciteit die omgekeerd evenredig is met de signaalbandbreedte. Het systeem is daarom uitgebreid met een recirculatie optie. Daarbij wordt de aanwezige overcapaciteit gebruikt om het aantal kanalen te vergroten tot een maximum van 40.000 kanalen. Hiertoe worden de ingangssignalen op de Nyquist snelheid (2B) gedurende (10/B).40 msec in een geheugen geschreven. De inhoud van dit geheugen wordt vervolgens N=10/B maal op maximale snelheid uitgelezen en naar de correlator gestuurd. De correlator berekent dan 10/B maal het aantal punten van de kruis-correlatiefunktie. De beschikbare kanalen worden verdeeld over nominaal 40 antenne combinaties. Voor 10 MHz bandbreedte kan dan per antennecombinatie een 128-punts kruis-correlatiefunctie gemeten worden die na Fouriertransformatie elke 10 seconden een 64-punts complex spectrum oplevert waarvan de amplitude en de fase de ruimtelijke informatie bevatten. Deze ruimtelijke informatie wordt 12 uur lang gemeten en resulteert na een 2-dimensionale Fouriertransform in een 1024x1024 punts intensiteits verdeling van het waargenomen object voor ieder van de 64 frekwentie punten.

Het tweede systeem (zie Fig. 3.2.) wordt gebruikt voor breedband continuum waarnemingen. Het bestaat uit een 8-kanaals analoge filterbank voor ieder antennesignaal. Ieder filter bestrijkt een band van 10 MHz zodat met een totale bandbreedte B van 80 MHz gemeten kan worden.

De reden voor het opsplitsen in banden is enerzijds het verminderen van de gevoeligheid voor storing, anderzijds het voldoende laag houden van de klok-frequentie van het digitale deel. De banden zijn verstembaar om selectie van storingsvrije gebieden in het spectrum mogelijk te maken. Tussen de 8x28 filter uitgangen worden, na analoog-digitaal conversie, in een digitale correlator 8x256 complexe 2-bit producten gevormd met een kloksnelheid van 20 MHz die gedurende 10 seconden worden geïntegreerd alvorens ze verder worden verwerkt. Dit is dus (zie 2a) een filterbank met een digitale detector.

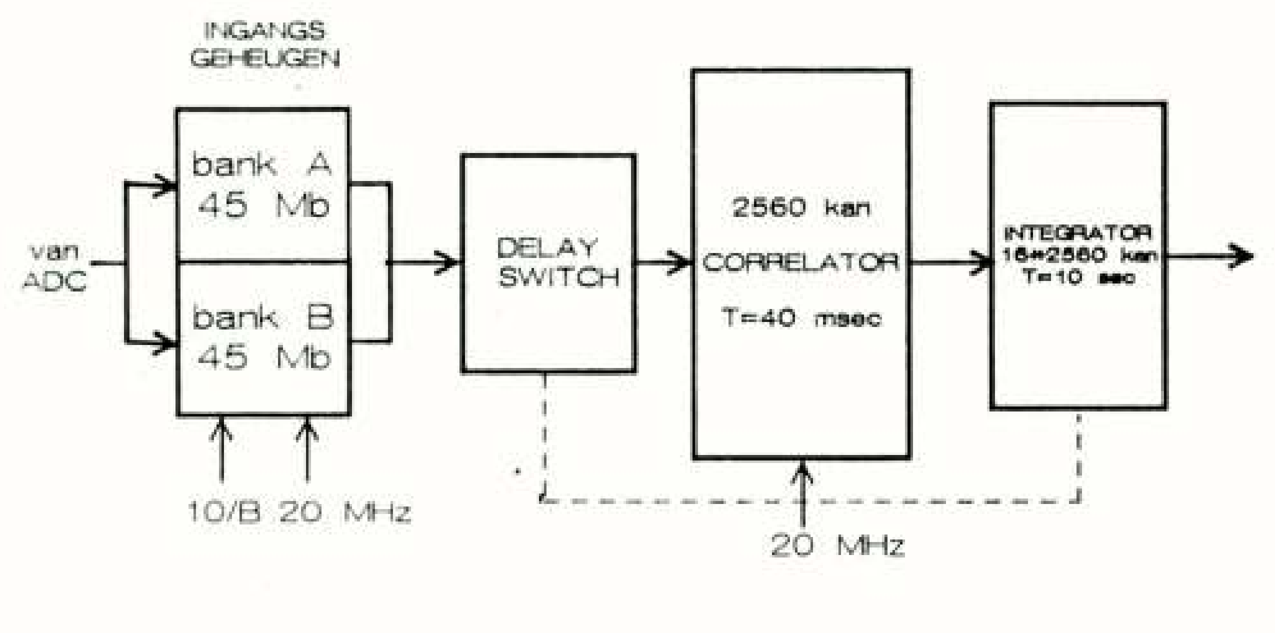


Fig. 3.1 WSRT spectraallijn correlator

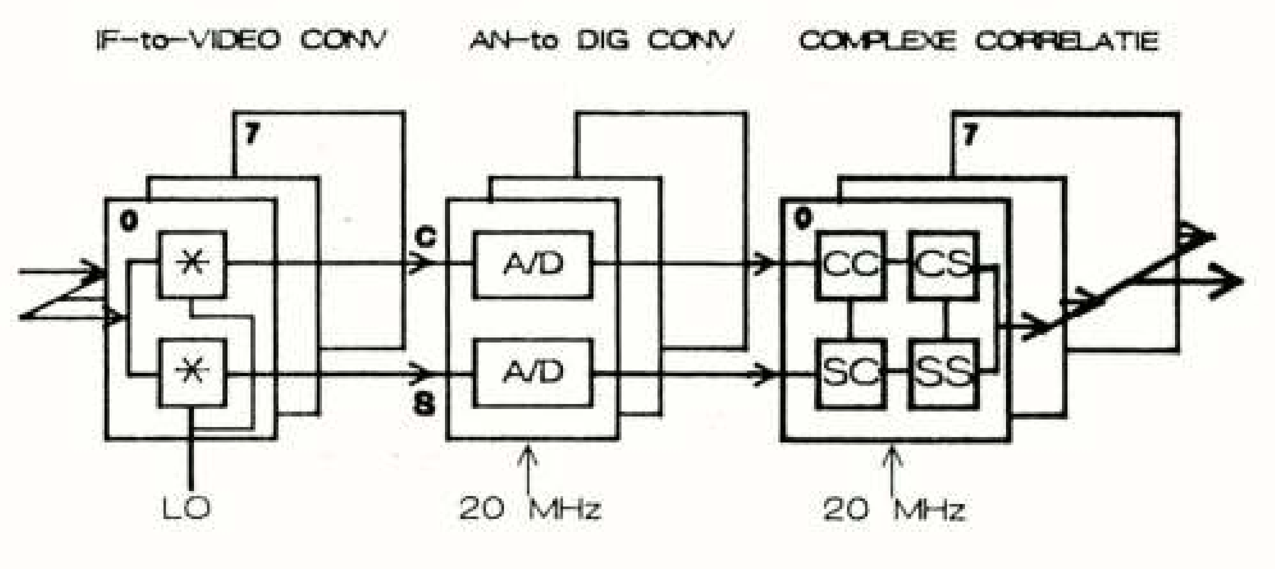


Fig 3.2 WSRT continuum correlator

De gedurende 12-uur verzamelde complexe producten bevatten de ruimtelijke informatie per band van 10 MHz. Het resultaat is dan equivalent aan dat van de hierboven genoemde spectrometer na de tijd-frekwentie transform. Het eindresultaat is, na de 2-dimensionale ruimtelijke Fouriertransform, echter één enkele intensiteitsverdeling waarin de informatie van de 8 frequentiebanden is gecombineerd.

4. Een hybride spectrometer voor de JCMT

Voor de JCMT is een spectrometer in aanbouw met een bandbreedte van 1 GHz. Om de in (2e) genoemde redenen is deze band opgesplitst in 8 banden van 125 MHz. In werkelijkheid gebruiken we echter subbanden met een 20 dB bandbreedte van 160 MHz die elkaar op de 3 dB punten overlappen om het combineren van de 8 subbanden tot één continue doorlaatband mogelijk te maken zonder dat

daarbij last wordt ondervonden van instabiliteiten en postdetectie ruis aan de rand van de doorlaatband. Bovendien is het mengschema zo gekozen dat eerste orde ongewenste mengproducten in het overlap gebied terecht komen.

De 160 MHz banden worden bemonsterd en gekwantiseerd door een 2-bits A/D omzetter met een effectieve snelheid van 320 MHz. Deze informatie wordt toegevoerd aan een digitale autocorrelator die na Fouriertransformatie een 128 punts spectrum produceert. Per 125 MHz band blijven er wegens de overlap effectief 100 punten over.

Realisatie van een correlator met deze kloksnelheid valt binnen de mogelijkheden van de huidige GaAs technologie maar is erg kostbaar, allereerst financieel en vervolgens in tijd en mankracht. Een goed alternatief wordt geboden door de toepassing van parallel verwerking met behulp van, op een lagere snelheid werkende, identieke bouwstenen. De uitgangssignalen van de A/D omzetters worden hiertoe via een serie-parallel

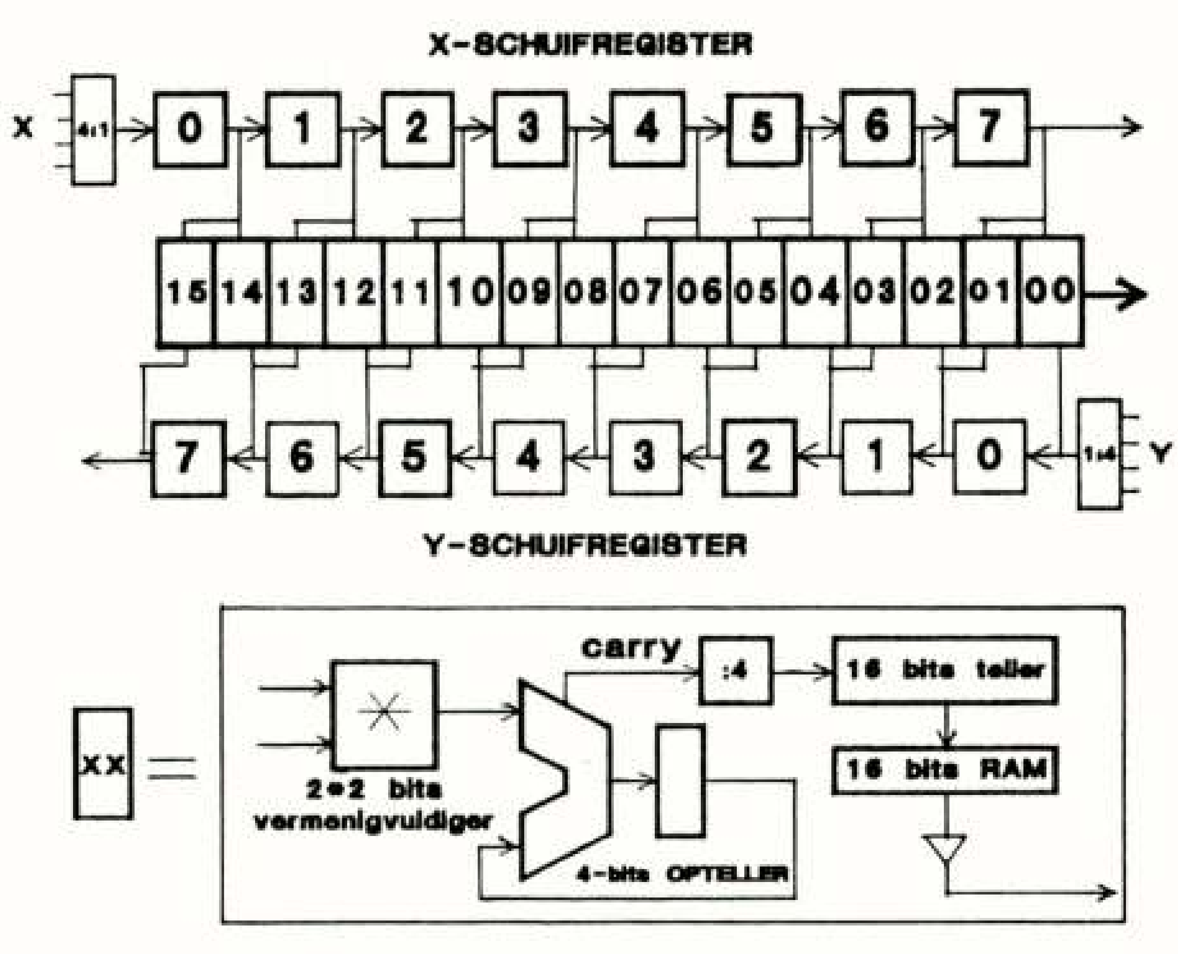


Fig. 5.1 De correlator cel

conversie in 8-bits woorden aan een correlator module toegevoerd. Een correlator module bestaat uit een 8x8 matrix van 64 correlator cellen die de 64 unieke correlaties tussen de bits van 2 opeenvolgende woorden maken met een kloksnelheid van 40 MHz. Elke correlator cel berekent 16 punten van de auto-correlatiefunctie. Per punt wordt iedere 25 nsec een 2x2 bits vermenigvuldiging, een 4 bits optelling en een schuifoperatie op de ingangssignalen uitgevoerd. Het fysische aantal punten per module is dus 1024. De inhoud van de 8 cellen met de parallelle informatie wordt na integratie bij elkaar opgeteld zodat er een correlatiefunctie met 128 punten overblijft.

5. De correlator cel

Uitgangspunt bij het ontwerp van de correlator cel was de wens om, gezien de ermee gemoeide kosten, de bouwsteen ook bruikbaar te maken voor de volgende generatie spectrometers. Daartoe werd een onderzoek gedaan naar de gemeenschappelijke kenmerken van de architecturen voor de verschillende toepassingsgebieden. Als resultaat hiervan werd een basis architectuur met een bijbehorende bouwsteen gedefinieerd waarmee de volgende generatie spectrometers gerealiseerd kan worden.

De eisen die daaruit voor de bouwsteen volgen zijn o.a.

- reconfiguratie moet mogelijk zijn om, gegeven een bepaalde correlator capaciteit, ruimtelijk oplossend vermogen uit te kunnen wisselen tegen spectraal oplossend vermogen en/of spectraal oplossend vermogen tegen bandbreedte.
- het moet mogelijk zijn om met een minimum aan externe componenten een maximum aan flexibiliteit te realiseren.
- een zo hoog mogelijke snelheid gekoppeld aan lage dissipatie.
- moet in serie geschakeld kunnen worden
- moet betaalbaar zijn.

Het resultaat [11] is een circuit met de volgende eigenschappen:

- de **kloksnelheid**: voor de gebruikte 1.5 μ CMOS technologie bleek een minimum kloksnelheid van 55 MHz mogelijk.
- de **vermenigvuldigingstabel**: een compromis tussen complexiteit en verlies in gevoeligheid. Een goed compromis bleek een gereduceerde 2x2 bit correlator te zijn die een gevoeligheid verlies geeft van 14%.
- het **aantal teller trappen**: De ruis op een N-bits teller is $-\sqrt{N}$ bits zodat alleen de meest significante bits uitgelezen behoeven worden. In ons geval gaat dit om een totaal van 22 bits waarvan meest significante 16 bits worden uitgelezen.
- **technologie**: CMOS gate array met 8000 cellen bleek het meest geschikt. Het bleek mogelijk om hierin 16 kanalen onder te brengen. (1 kanaal bestaat uit een schuifregistertrap, een 2x2 bits vermenigvuldiger, een 22-bits teller en een 16-bits buffer).

Het blokschema (zie Fig. 5.1.) toont een bidirectioneel schuifregister dat is toegepast om de distributie van de signalen naar de vermenigvuldigers eenvoudig te houden. De vermenigvuldiger is uitgevoerd in 3 synchrone trappen, die een asynchrone teller sturen. Voor configuratie selectie zijn 1:4 selectors aangebracht.

De correlator sectie van de hybride spectrometer bevat 1024 cellen die ondergebracht zijn op 16-tal 10-laags printplaten. Een printplaat bevat de bovengenoemde 8x8 matrix van 64 cellen. De 16384 16-bits celinhouden worden elke 13 msec uitgelezen en verder opgeteld in een integrator. De integrator wordt op zijn beurt na 10 sec uitgelezen door de besturingscomputer die het resultaat verder verwerkt tot een interpreteerbaar spectrum.

Referenties

[1] Bos, A., 1985, On instrumental effects in spectralline synthesis observations, Thesis, Leiden University.

[2] Penfield, H., 1976, Methods of experimental Physics, 12B, 266-279, New York, Academic Press.

[3] O'Sullivan, J.D., 1977, Tijdschrift van het NERG, 42, 1/2, 5-13.

[4] Weinreb, S., 1963, M.I.T., Technical Report 412.

[5] Cooper, B.F.C., 1970, Aust. J. Phys. 23, 521-527.

[6] Chikada, Y. et al., 1983, Indirect Imaging, 287-404, Ed. Roberts, J.A., Cambridge University Press, Cambridge.

[7] Masson, C., 1982, Astron. Astrophys., 144, 270-274.

[8] Bos, A., 1977, Tijdschrift van het NERG, 42, 1/2, 15-19.

[9] Bos, A., Raimond, E., van Someren Greve, H.W., 1981, Astron. Astrophys., 98, 251-259.

[10] O'Sullivan, J.D., 1983, Indirect Imaging, 405-413, Ed., Roberts, J.A., Cambridge University Press, Cambridge.

[11] Bos, A., 1989, NFRA Internal Technical Report 176.

Voordracht gehouden tijdens de 368e werkvergadering.

SCHOTTKY-MIXER ONTWIKKELING VOOR HET SUB-MILLIMETER GOLFLENGTE GEBIED

Ir. E.E.M. Woestenburg
Stichting ASTRON
Radiosterrenwacht Dwingeloo
Postbus 2, 7990 AA DWINGELOO

Schottky-mixer development for sub-millimeter wavelength. The design and modelling of Schottky-diode mixers for sub-millimeter wavelength is presented. Calculated performance based on the mixer model agrees well with measured results on waveguide mixers realised for frequency bands around 345 GHz and 470 GHz.

Inleiding

Schottky-mixers voor sub-mm golflengten vinden toepassing als element voor frequentie konversie in ontvangers voor radio-astronomisch gebruik. Als gevolg van de absorbtie door de atmosfeer van signalen in het sub-mm golflengtegebied is waarneming van astronomische objecten bij deze golflengten alleen mogelijk vanaf hoog gelegen observatoria, zoals de James Clerk Maxwell Telescope (JCMT) op de 4200 m hoge berg Mauna Kea op het eiland Hawaii. Voor deze teleskoop ontwikkelt en bouwt ASTRON in internationaal verband ontvangerapparatuur, als één van de ondersteunende laboratoria. De JCMT en de ervoor ontwikkelde ontvangers werden reeds uitvoerig beschreven in een eerder artikel (Van Ardenne, 1987), waarnaar voor meer achtergrond-informatie wordt verwezen. Dit artikel beperkt zich tot een kort overzicht van de huidige technische stand van zaken met betrekking tot de teleskoop en ontvangers voor lijnstraling. Daarnaast wordt ingegaan op het ontwerpen van Schottky-mixers voor frequenties boven 300 GHz. Tenslotte komen enkele andere ontwikkelingen aan de orde.

Huidige situatie

Sinds de officiële openstelling van de JCMT in april 1987 is de nauwkeurigheid van het teleskoopoppervlak, van de aanvankelijk bereikte waarde van 50 µ rms, stapsgewijze verbeterd tot 30 µ rms. Daarmee ligt de oppervlaktenauwkeurigheid van de teleskoop met een diameter van 15 m ruimschoots binnen de bij het ontwerp als doel gestelde 35 µ rms. Verwacht wordt dat in de nabije toekomst nog enige verbetering mogelijk is. Gebruik voor golflengten tot 0.3 mm ligt hierdoor binnen de mogelijkheden.

Wat de ontvangers betreft is tot nu toe voornamelijk gebruik gemaakt van Schottky-mixer ontvangers voor frequenties rond resp. 230 GHz en 345 GHz. Daarnaast zijn incidenteel testen en waarnemingen gedaan met "gast ontvangers", met name een Schottky-mixer ontvanger voor frequenties rond 800 GHz en ontvangers die gebruik maken van supergeleidende (SIS) detectoren voor 230 GHz en 345 GHz. De grotere gevoeligheid van ontvangers met dit type detector zal ertoe leiden dat op termijn van een jaar de bestaande (Schottky) ontvangers voor de lagere frequenties zullen worden vervangen door ontvangers met SIS-detectoren. Ook voor hogere frequenties bestaat de verwachting dat SIS-ontvangers goede resultaten zullen opleveren en het huidige onderzoek concentreert zich dan ook voornamelijk op de ontwikkeling van SIS-juncties voor deze toepassing. Daarnaast is een enige jaren geleden gestarte ontwikkeling voor Schottky-mixers bij 470 GHz recent afgerond en het resultaat op korte termijn toepasbaar.

Principe Schottky-mixer

Aangezien voor het van belang zijnde frequentiegebied geen versterkerelementen beschikbaar zijn, wordt het inkomende signaal m.b.v. een lokaal oscillatorsignaal gemengd naar een lagere frequentie, waar versterking wel mogelijk is. Figuur 1 toont het principe schema voor heterodyne menging met een sub-millimeter mixer. Voor deze heterodyne menging wordt gebruik gemaakt van een niet lineair element in de vorm van een Schottky diode. Deze diode bestaat uit een metaal/halfgeleider overgang met een stroom-spanningskarakteristiek volgens:

$$i = i_0 \left(\exp\left[e\, V_j / k\eta\, T_0\right] - 1\right) \qquad (1),$$

waarin i en Vj resp. de stroom door en de spanning over de diode zijn, k de konstante van Boltzmann is en η een kwaliteitsfaktor voor de diode (1.05-1.2). T_0 is de diode temperatuur in K, i_0 is de verzadigingsstroom (reverse bias). De kwadratische komponent van de stroom, via een reeksontwikkeling uit (1) te verkrijgen, is verantwoordelijk voor menging van de toegevoerde signalen. Gebruikmakend van harmonische ingangssignalen $a\cos \omega_{LO}t + b\cos \omega_S t$ volgt m.b.v. enkele simpele goniometrische bewerkingen dat de stroom door de diode komponenten op verschil- en somfrequenties van de ingangssignalen bevat: $\cos(\omega_{LO} \pm \omega_S)t$. De verschilfrequentie wordt uitgefilterd en naar een middenfrequent versterker gevoerd.

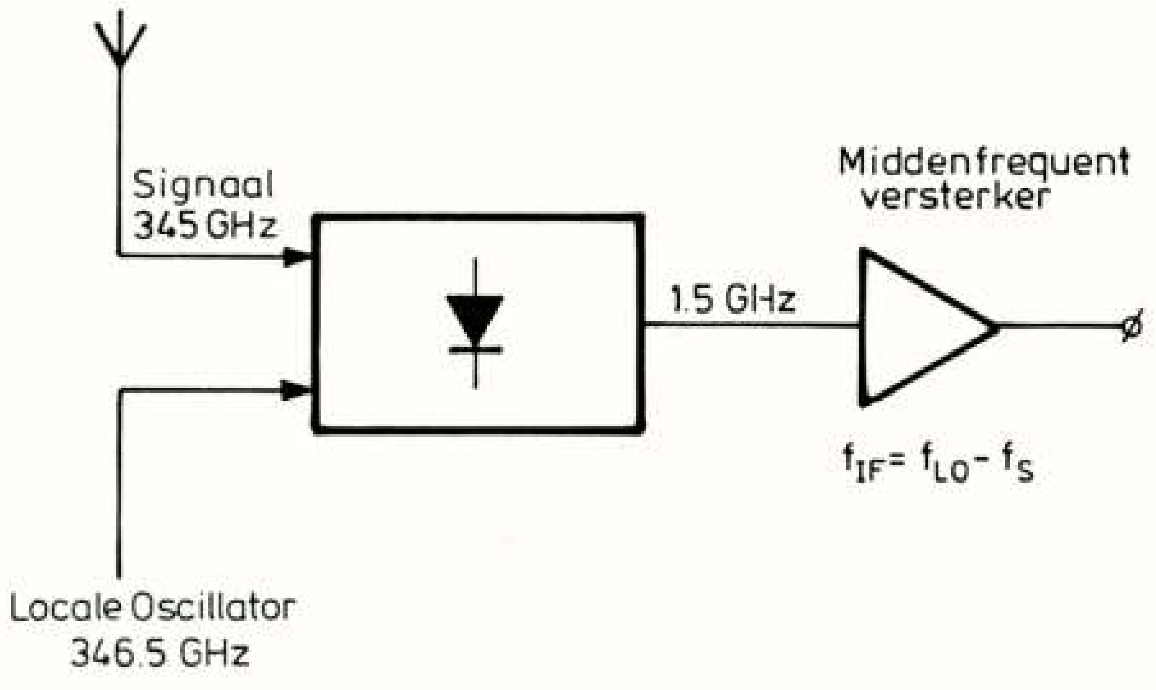


Fig. 1: Principeschema voor heterodyne menging met de sub-mm mixer. Links is de antennesektie aangegeven die het signaal op 345 GHz opvangt en naar de mixer leidt. Er wordt eveneens een oscillatorsignaal in de mixer geïnjekteerd om aan de uitgangspoort het middenfrequent-signaal, op de verschilfrequentie, te verkrijgen. De mixer wordt gevolgd door een lageruis FET-versterker.

De Schottky diode

Diodes voor toepassing bij sub-millimeter golflengten worden gekenmerkt door kleine afmetingen. Hieraan ligt ten grondslag dat een zeer lage diode junctie capaciteit (enkele femto Farad, 10^{-15}F) en kleine parasitaire elementen noodzakelijk zijn om een goede aanpassing aan de microgolfomgeving mogelijk te maken. Deze diodes worden gemaakt bij slechts enkele universitaire

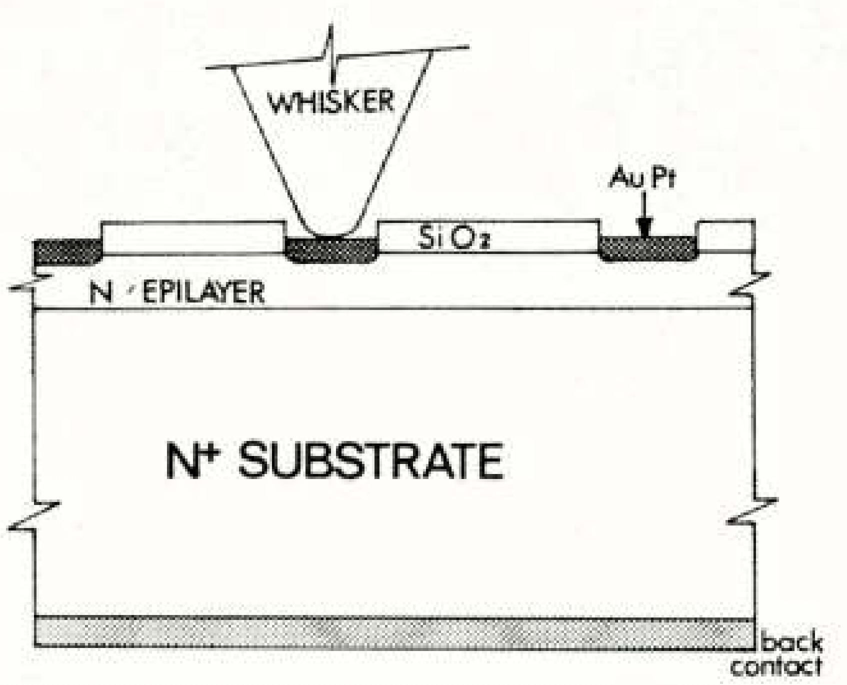


Fig. 2: Doorsnede van een gedeelte van een mm diode-chip. De onderzijde van de chip wordt op het uiteinde van de coaxiale struktuur in de mixer gesoldeerd. De punt van whisker wordt op één van de anode eilandjes aan de bovenzijde van de chip gedrukt.

instituten in de USA en Europa en bestaan uit een honingraatsstruktuur van 100 μ x 100 μ met daarop enkele tientallen halfgeleidereilanden (anodes Ø 2 μ) als mogelijke diodekontakten. Figuur 2 laat een gedeelte zien van een millimeter-diode chip. Het diodekontakt dient, na montage van de diode in een mixer, m.b.v. een dunne gebogen draad op één van de anodevlakjes gemaakt te worden. Deze techniek vertoont veel overeenkomst met de kristaldetektor uit de beginjaren van de radiotechniek waar een puntkontakt op een kristal een diode-overgang vormde. De betreffende dunne draad, de z.g. whisker, is van fosforbrons en heeft een lengte van ongeveer 100 μ bij een diameter van 7 μ. De handelingen om tot het maken van een diodekontakt te komen, zoals het in de juiste vorm buigen van de whisker en het vormen van het kontakt, vinden plaats m.b.v. micromanipulatoren onder een microscoop.

De mixer

Voor het ontwerp van de Schottky mixer is gekozen voor een uitvoering in rechthoekige golfpijp (0.65 mm x 0.20 mm voor 345 GHz) met een coaxiale uitkoppeling van het middenfrequent signaal bij 1.5 GHz. Het LO- en RF-signaal wordt ingekoppeld via een rechthoekige z.g. pyramidale hoorn waarvoor een diëlektrische lens is opgesteld die het inkomende signaal focusseert. De microgolfhoorn is via een getaperde golfpijpsektie verbonden met de golfpijp waarin de diode is gemonteerd. De afmetingen van de golfpijp worden bepaald door de dominante golfpijp mode ($TE_{1,0}$). De maximum diameter van de coaxiale struktuur wordt gegeven door de formule:

$$D+d < 190/f \qquad (2),$$

waarin d en D resp. de binnen- en buitendiameter van de coaxiale geleider met lucht als diëlectricum zijn en f de frequentie in GHz is. Uit (2) volgt dat $D \leq 0.28$ mm voor het frequentiegebied rond 345 GHz.
In de coaxiale geleider wordt een laagdoorlaatfilter opgenomen om de middenfrequentuitgang te isoleren van de golfpijp bij 345 GHz. Om dit filter te realiseren wordt een 5-sektie filter toegepast met een kleinste diameter van 0.1 mm. Figuur 3 toont een doorsnede van een 345 GHz golfpijp mixer zoals in gebruik bij de JCMT. De belangrijkste afmetingen zijn in de figuur aangegeven.
Eén van de grootste problemen bij de realisatie van een dergelijke golfpijp- mixer is de fabrikage van onderdelen met de genoemde kleine afmetingen. Hiervoor zijn speciale technieken, hulpinstrumenten en de bijzondere vaardigheden van mechanische en elektronische technici noodzakelijk. De coaxiale filterstruktuur, met een totale lengte van 0.8 mm, wordt op een precisie draaibank gedraaid, waarbij het gebruik van een microscoop onontbeerlijk is. De golfpijponderdelen worden gerealiseerd door gebruik te maken van de

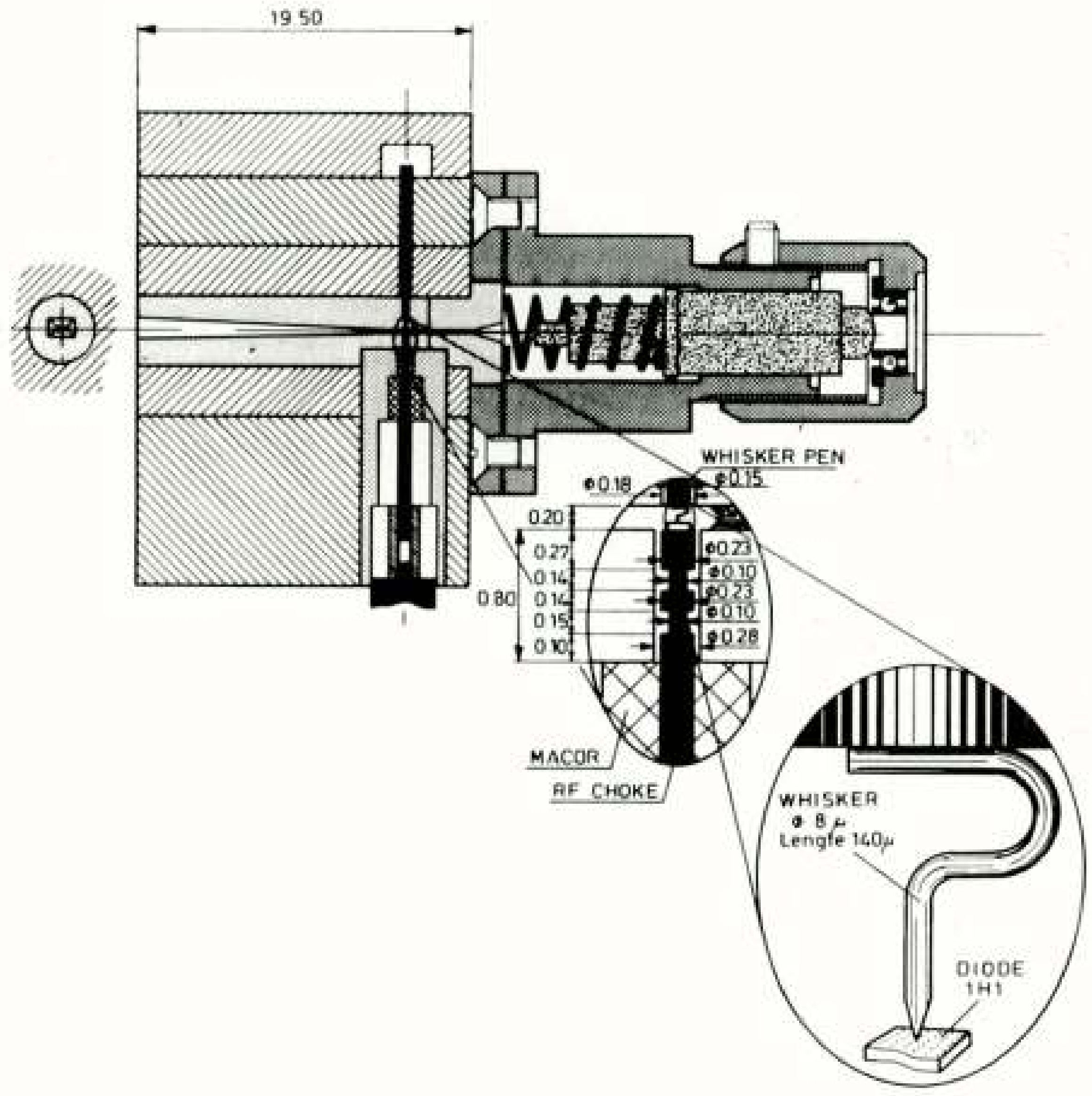


Fig. 3: Doorsnede van een 345 GHz mixer. Links is de microgolfhoorn als signaal- en LO-ingang van de golfpijp mixer aangegeven. De uitvergroting toont de plaats waar de diode in de golfpijp is bevestigd met details van de coaxiale struktuur, whisker en backshort (maten in mm). De uitkoppeling van het IF-signaal bij 1.5 GHz vindt plaats via een 50 Ω SMA connector.

elektroformeertechniek. Speciale aandacht wordt bij de materiaalkeuze en konstruktie besteed aan cryogene eigenschappen voor het gebruik van de mixer bij temperaturen rond 20 K. Koeling tot cryogene temperatuur is noodzakelijk om zo gunstig mogelijke ruiseigenschappen van de mixer te verkrijgen en wordt bereikt m.b.v. een gesloten koelsysteem met gasvormig helium als koelmedium.

Voor het optimaliseren van de eigenschappen van de mixer voor de gewenste frequentieband wordt gebruik gemaakt van een beweegbare kortsluitzuiger (backshort) in de golfpijp en de lengte van de whiskerdraad waarvan de zelfinduktie gebruikt wordt om de diodecapaciteit uit te stemmen.

Mixer model

Bij het microgolftechnische ontwerp van de mixer dient gelet te worden op een juiste aanpassing van de diode-impedantie aan de omgeving (embedding impedantie). Een representatie van diode en omgeving in een model is essentieel om inzicht te krijgen in de invloed van de diode eigenschappen en embedding impedantie op, en het voorspellen van, de eigenschappen van de mixer. Figuur 4 geeft een vervangingsschema van de diode en de golfpijp-omgeving dat gebruikt is voor berekeningen van de eigenschappen van de mixer. Voorwaarde voor een goede aanpassing is dat de diode impedantie gelijk is aan de toegevoegd complexe waarde van de embedding impedantie:

$Z_d = Z_e^*$.

Voor de diode conductantie en capaciteit in fig. 4 gelden de volgende formules:

$$g(V) = \frac{qV}{\eta kT} I(V) \qquad (3)$$

$$C(V) = C_{jo}/(1-V/\emptyset_{bi})^{0.5} \qquad (4)$$

In (3) en (4) is V de spanning over de diode; verder geldt dat C_{jo} de diodecapaciteit is zonder spanning over de diode en is $\emptyset_{bi}$ een spanning die afhankelijk is van de materiaaleigenschappen.

M.b.v. het gegeven vervangingsschema is een rekenmodel gemaakt van de mixer. In het model wordt een "groot signaal" analyse uitgevoerd teneinde de spanningsgolfvormen over de diode te bepalen t.g.v. het LO-signaal. Deze golfvormen bepalen de niet-lineaire junctiestroom en capaciteit en daarmee de "klein signaal" golfvormen voor de junctie conductantie en capaciteit. Bij de analyse wordt ervan uitgegaan dat het inkomende RF-signaal verwaarloosbaar klein is t.o.v. het LO-signaal. In de analyse met de z.g. harmonic balance techniek wordt verder de embedding impedantie betrokken op de LO-frequentie en een aantal harmonischen daarvan.

De essentie van de harmonic balance techniek is het vinden van:

- een spanning en stroom in het tijddomein die voldoen aan de diodevergelijking.
- een oplossing in het frequentiedomein die voldoet aan de vergelijkingen van het externe circuit.

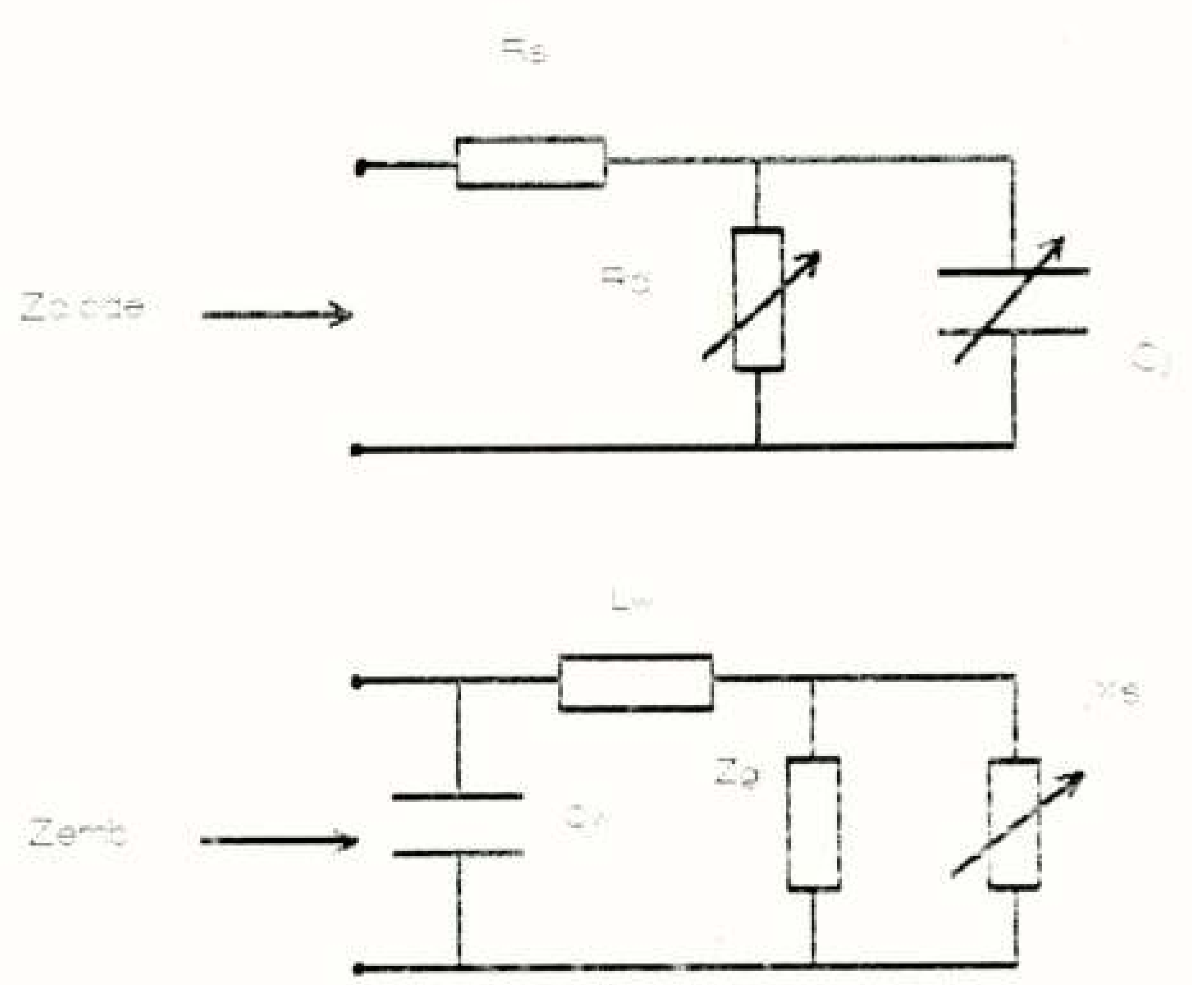


Fig. 4: Vervangingsschema's van de diode en microgolf omgeving, zoals gebruikt in een model voor de berekening van de eigenschappen van de mixer.

Beide oplossingen dienen, via hun Fourier reeks representatie, aan elkaar gelijk te zijn.
Het proces start door een aanname te doen voor de oplossing in het tijddomein en deze in te vullen in de diodevergelijking. Vervolgens wordt na Fourier transformatie gekeken of de oplossing voldoet aan de externe circuit vergelijkingen. Desgewenst wordt de oplossing bijgesteld totdat convergentie ervan is verkregen.
Frequentie menging in de diode ontstaat doordat de "klein signaal" junctie conductantie en capaciteit harmonisch variëren met het LO signaal. Met de gevonden golfvormen en de formules (3) en (4) worden nu de conversie eigenschappen van de diode nader onderzocht.
Het resultaat van de analyse is een conversie matrix voor de diode, waaruit met de gegevens van het embedding circuit mixereigenschappen zoals conversieverlies en in- en uitgangsimpedantie berekend kunnen worden.
Naast het conversieverlies is, met name voor de genoemde astronomische toepassingen, de eigen ruisbijdrage van de mixer van groot belang. Voor de bepaling van het ruisgedrag van de mixer wordt in het model rekening gehouden met twee ruismechanismen:

- shot noise door gedetekteerde LO-stroom, weergegeven door $\eta T_0/2$ (T_0-diode temperatuur, η-kwaliteitsfaktor)
- thermische (Johnson) ruis met als beschikbaar ruisvermogen P=kTB (k-konstante van Boltzmann, B bandbreedte in Hz).

Van belang bij de ruisanalyse zijn de correlaties tussen shot noise komponenten op diverse mengfrequenties, die bijdragen tot ruis op de IF-frequentie. De gebruikte ruisanalyse is gebaseerd op het werk van Held en Kerr (1978) en maakt gebruik van correlatiematrices voor shot noise en thermische ruis.

Mixer resultaten

Een op basis van de analyse ontwikkeld PC-programma berekent als funktie van de frequentie o.a. conversieverlies en ruistemperatuur van de mixer bij afstemming van de backshort voor minimale ruis. Als parameters kunnen eigenschappen van de diode en het embedding cricuit worden ingevoerd.
Op deze manier is het mogelijk de invloed van diode parameters, zoals junctie capaciteit en serie weerstand, op het gedrag van de mixer te bepalen en diodes op basis hiervan voor een bepaald frequentiegebied te selekteren.
De lengte van de whisker wordt in de praktijk, naast de beweegbare backshort, als parameter gebruikt om het mixergedrag te optimaliseren. In figuur 5 is de berekende invloed van de whiskerlengte op het ruisgedrag weergegeven als funktie van de frequentie rond 470 GHz. Voor een breedbandige responsie blijkt een whiskerlengte rond 50 μ optimaal. Het getoond gedrag als funktie van de lengte van de whisker is karakteristiek voor dit type mixer. Voor ontwerpen in de 345 GHz- en 470 GHz-band blijken berekende resultaten redelijk overeen te stemmen

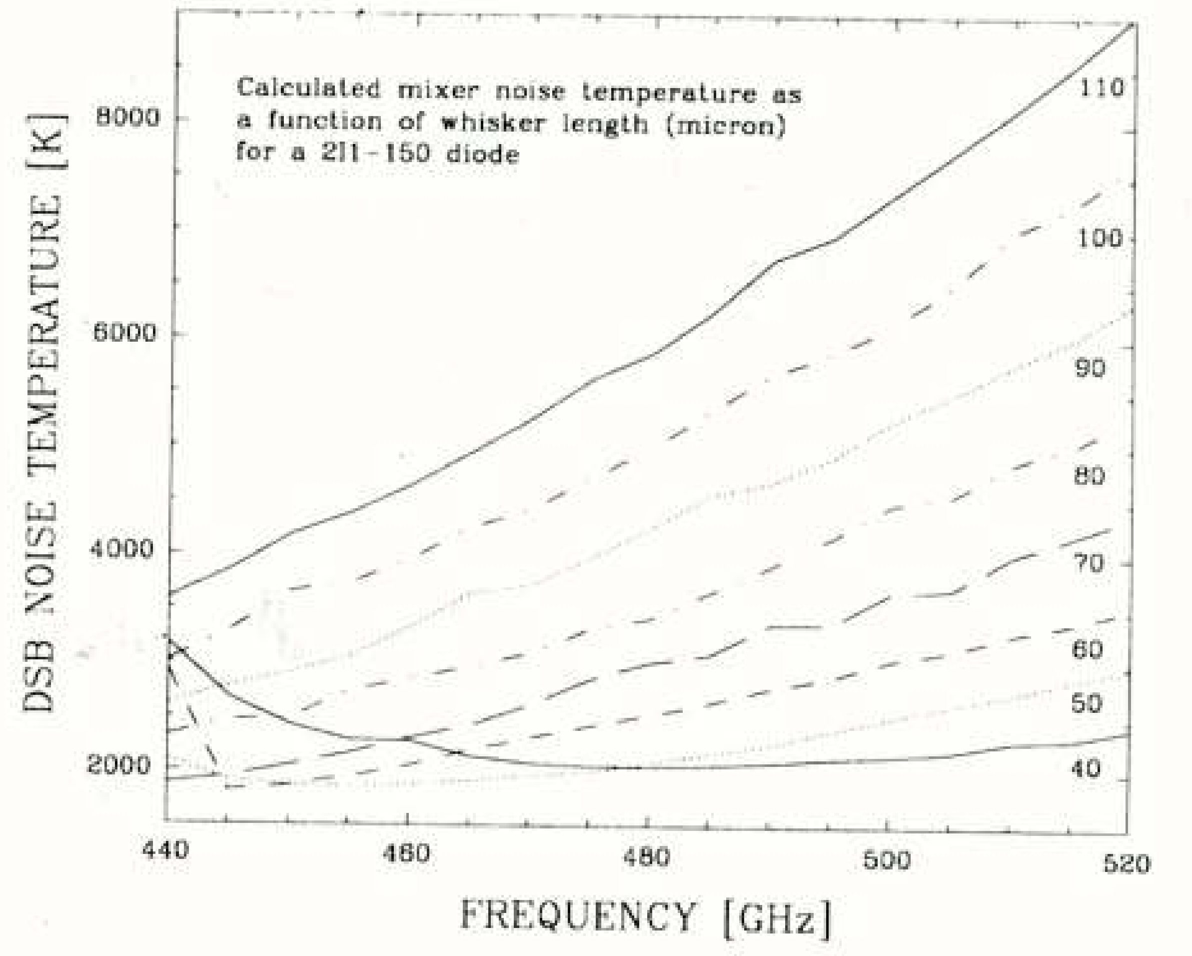


Fig. 5: Volgens het model berekende invloed van de lengte van de whisker op de ruiseigenschappen van een mixer ontwerp rond 470 GHz.

met gemeten eigenschappen. Fig. 6 laat dit zien aan de hand van een berekend en gemeten resultaat bij kamertemperatuur voor een 470 GHz mixer met een whiskerlengte van 70 μ. Overigens moet worden opgemerkt dat een juiste bepaling van de, in het model gebruikte, diodeparameters van essentieel belang is om tot een goede overeenstemming tussen model en praktijk te komen.

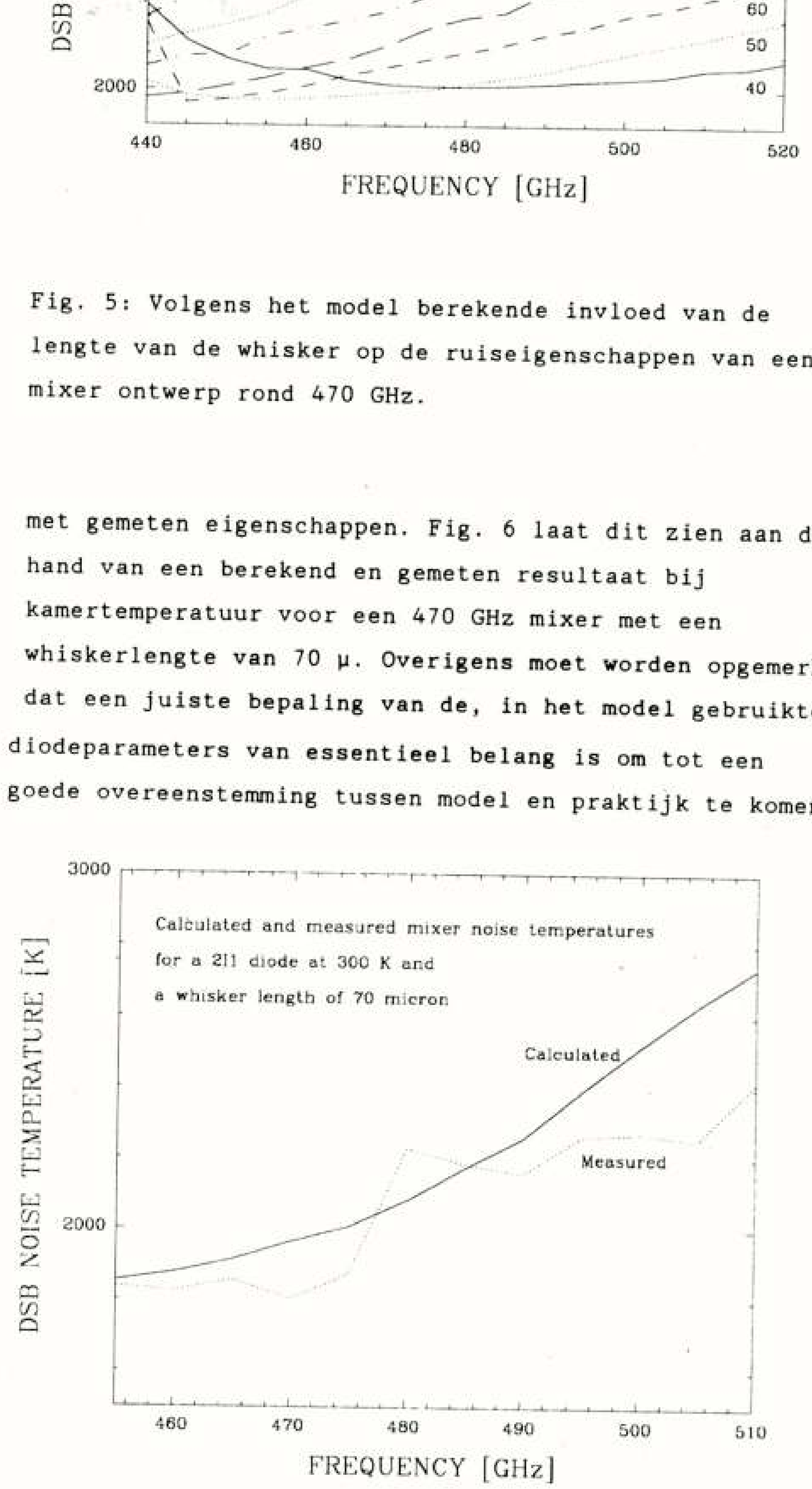


Fig. 6: Berekend en gemeten resultaat voor een mixer werkend bij kamer-temperatuur en een frequentie van 470 GHz. De whiskerlengte bedraagt 70 μ. Uit de figuur blijkt een redelijke overeenstemming tussen berekende en gemeten waarden.

Om de ruiseigenschappen te verbeteren wordt de mixer gekoeld tot cryogene temperaturen (rond 20 K) m.b.v. een gesloten koelsysteem. Met de temperatuurverlaging neemt

zowel de thermische ruisbijdrage (evenredig met de temperatuur) als de shot noise af. De ruisverlaging die op deze manier haalbaar is, blijkt uit figuur 7 waarin

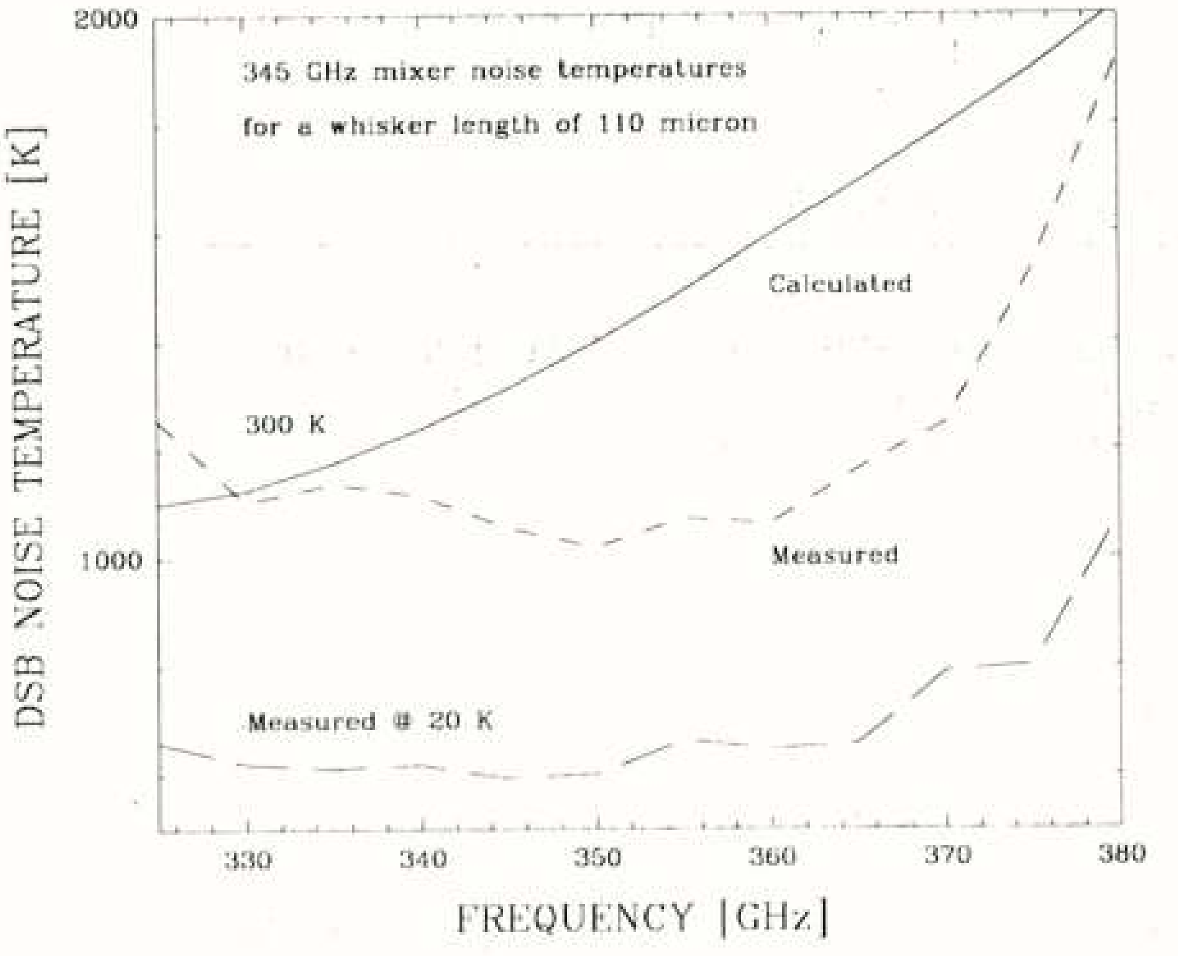


Fig. 7: Berekend en gemeten ruisgedrag van de in Hawaii gebruikte 345 GHz mixer. De verbetering van de ruistemperatuur door afkoelen tot 20 K is signifikant (zie tekst). De lengte van de whisker bedraagt voor deze mixer 110 μ.

de ruistemperaturen als funktie van de frequentie worden weergegeven bij kamertemperatuur (berekende en gemeten waarden) en gekoeld tot 20 K (gemeten waarden) voor één van de mixers die in de 345 GHz ontvanger op Hawaii in gebruik is. Het getoonde resultaat voor de gekoelde mixer is inclusief de ruisbijdrage van de gekoelde IF-versterker en vertegenwoordigt in feite de ontvanger ruistemperatuur. De verbetering in ruistemperatuur van de mixer op zich bedraagt een faktor 2,5 tot 3 en is daarmee groter dan de faktor 2 die uit figuur 7 zou volgen.

Nieuwe ontwikkelingen

De praktische realisatie van golfpijpmixers volgens het hier gepresenteerde ontwerp blijkt tot een frequentie van 500 GHz mogelijk. Voor hogere frequenties wordt de grens van wat met de genoemde technieken haalbaar is, overschreden.

Hoewel theoretische simulatie bij 690 GHz bruikbare mixer eigenschappen voorspelt en ook geschikte diodes verkrijgbaar zijn, ontstaan er problemen m.b.t. de afmetingen van de golfpijp, coaxstruktuur, lengte van de whisker en de beschikbaarheid van voldoende LO-vermogen. Voor de praktische realisatie zullen andere uitvoeringsvormen van een mixer onderzocht moeten worden. Hierbij kan gedacht worden aan de toepassing van eenvoudiger te realiseren, circulaire, golfpijpen en planaire of quasi-optische strukturen. Om het tekort aan LO-vermogen te kompenseren is het gebruik van supergeleidende strukturen (SIS-juncties) als niet-lineair element mogelijk. Deze hebben daarnaast het voordeel dat potentieel een grotere gevoeligheid (lagere eigen ruisbijdrage) kan worden bereikt dan met Schottky diodes en dat zij door hun planaire struktuur eenvoudiger te integreren zijn in meer-element mixer arrays. Het belang hiervan is reeds door Van Ardenne (1987) aangegeven. In de ondersteunende laboratoria voor de JCMT concentreert men zich daarom voor de ontwikkeling en bouw van nieuwe ontvangers voornamelijk op de toepassing van SIS-juncties voor enkelvoudige en meer-bundelsystemen.

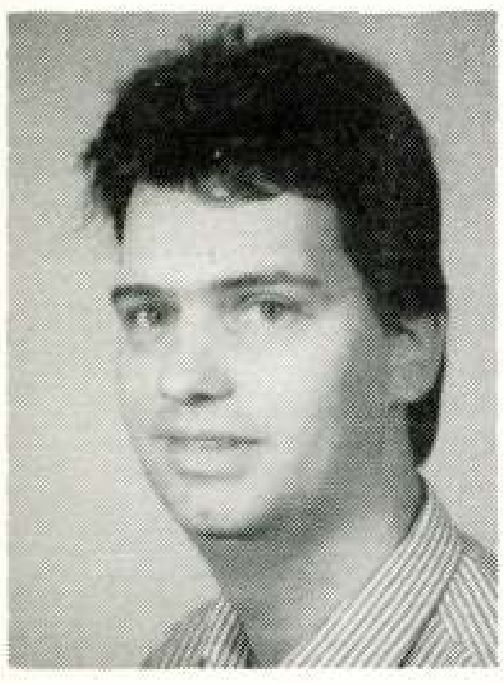
Referenties:

Van Ardenne, A. (1987); Ontvanger ontwikkelingen voor de James Clerk Maxwell Telescope (JCMT), Tijdschrift NERG deel 52, nr. 4/5/6, blz. 121-130.

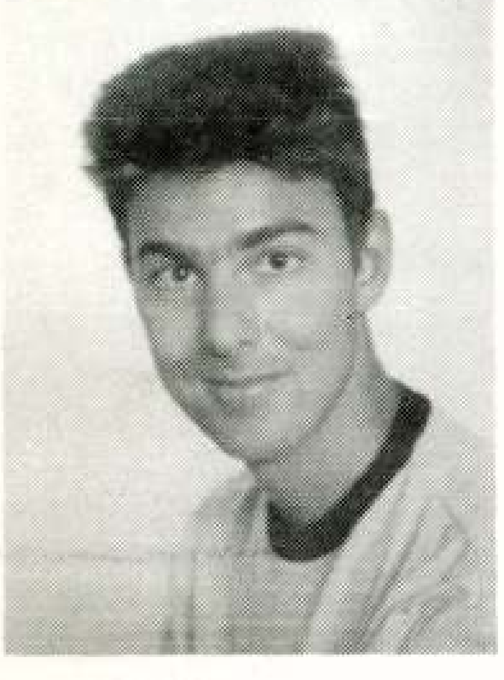
Held, D.N., Kerr, A.R.; Conversion Loss and Noise of Microwave and Millimeter-wave Mixers: Part 1 - Theory; Part 2 - Experiment, IEEE Trans. Microwave Theory Tech., MTT-26, 1978, p. 49.

Voordracht gehouden tijdens de 368e werkvergadering.

NEDERLANDS ELEKTRONICA- EN RADIOGENOOTSCHAP
(376ste werkvergadering)
IEEE BENELUX SECTIE
AUDIO ENGINEERING SOCIETY



DR. IR. A. BOSSCHE



IR. H. C. J. M. VAN GESTEL

UITNODIGING

voor de lezingendag op **donderdag 5 april 1990** in het **PTT-vergadercentrum, Burg. Fockema Andreaelaan 15, Utrecht.**
THEMA: ELEKTRONICA EN BEDRIJFSZEKERHEID.

PROGRAMMA:

11.00 uur:	Ontvangst en koffie.
11.30 uur:	**DR. IR. A. BOSSCHE,** (Technische Universiteit Delft, Faculteit der Elektrotechniek): "BEDRIJFSZEKERHEID, LUXE OF NOODZAAK?"
12.00 uur:	**IR. H. C. J. M. VAN GESTEL,** (Technische Universiteit Delft, Faculteit der Elektrotechniek): BEDRIJFSZEKERHEIDSONDERZOEK AAN IC PACKAGING.
12.30 uur:	Lunch.
14.00 uur:	**DR. J. J. M. KOOMEN,** (Philips Components, Advanced Development and Manufacturing Centre (ADMC), Waalre): "BETROUWBAARHEIDSASPECTEN BIJ IC-FABRICAGE."
14.45 uur:	Thee.
15.15 uur:	**IR. D. BLOEMBERGEN,** (Philips Consumer Electronics, Component Investigation and Reliability Group (CIRG), Eindhoven): "CONSUMENTENELEKTRONICA: NOOIT STUK?"
16.00 uur:	Sluiting.

De lezingen worden voorafgegaan door de jaarlijkse algemene vergadering van het NERG.
Aanmelding voor de lezingen dient te geschieden vóór 25 maart door middel van de aangehechte kaart **gefrankeerd met 55 cent.** Reservering voor de lunch vindt slechts plaats als **vóór 30 maart** een bedrag van **f 15,00** is ontvangen op de postrekening 94746 t.n.v. Penningmeester NERG, Leidschendam, onder vermelding van "E&B"-, NERG-, IEEE-, AES-leden en studenten hebben gratis toegang.
Tevens kunnen studenten de helft van de vervoerskosten vergoed krijgen (openbaar vervoer of anders, gezamelijk, mits goedkoper).
Niet-leden dienen een entree-prijs van f 15,00 te betalen. Deelnemers dienen **de uitnodigingskaart mee te nemen** en op verzoek te tonen bij de toegang tot het gebouw.
Het PTT-vergadercentrum is per bus vanaf het Centraal Station bereikbaar met de bus richting Rubenslaan (lijn 3 of 42). Per auto kan het vergadercentrum bereikt worden door op de autosnelweg Den Haag-Arnhem de afslag richting Amersfoort te nemen. Op deze weg neemt u daarna de afslag "de Uithof" en rijdt u onder het viaduct (linksaf) richting Utrecht-Centrum. Na enkele honderden meters ziet u aan de rechterzijde van de weg het gebouw van het Telecommunicatie-district.
NERG-leden die de algemene vergadering wensen bij te wonen, dienen dit aan te geven op de aangehechte kaart. Tevens dient te worden aangegeven of men de jaarstukken wenst te ontvangen.

Namens de samenwerkenden verenigingen,
IR. N. H. G. BAKEN, NERG.
070-3326482.

Leidschendam, maart 1990.

BEDRIJFSZEKERHEID, LUXE OF NOODZAAK

dr. ir. A.Bossche
Technische Universiteit Delft
Faculteit der Elektrotechniek

Reliability Engineering, luxury or necessity? Manufacturers and (professional) users show a growing interest in the reliability of their products and systems. This paper discusses the reasons for this interest and shows that reliability engineering is an economically sound activity which strengthens the competitive position.

Fabrikanten en (professionele) gebruikers raken steeds meer geinteresseerd in de bedrijfszekerheid van hun producten en systemen. Het is dan ook interessant om te bekijken waar deze belangstelling op gebaseerd is. Oftewel, zijn er gegronde redenen aan te wijzen welke bedrijfszekerheidstechniek tot een zinvolle bezigheid maken of betreft het slechts een voorbijgaand modeverschijnsel?
Voordat we deze vraag kunnen beantwoorden dienen we ons af te vragen wat onder het begrip 'bedrijfszekerheid' verstaan wordt.

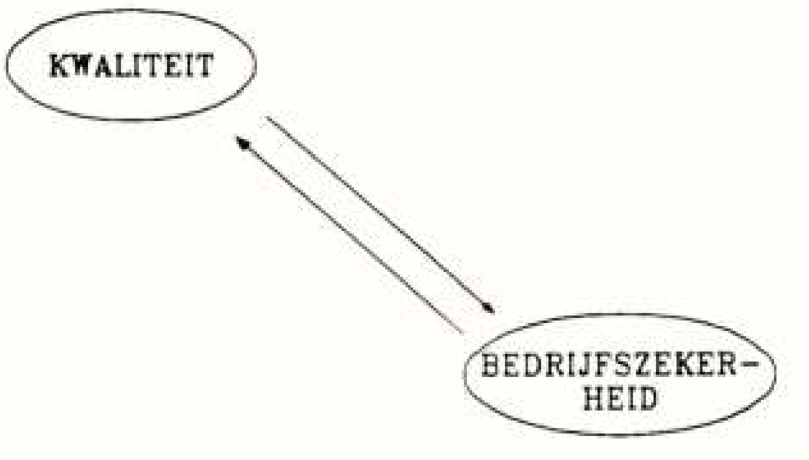


Fig. 1: De relatie tussen kwaliteit en bedrijfszekerheid; de bedrijfszekerheid geeft aan hoe lang, naar verwachting, de kwaliteit behouden zal blijven. Hoe langer de kwaliteit gehandhaafd kan worden hoe beter de bedrijfszekerheid.

Kwaliteit en bedrijfszekerheid

Wanneer een fabrikant een product vervaardigd of een gebruiker zichzelf een product aanschaft dan zullen deze ongetwijfeld een aantal eisen stellen waaraan het product moet voldoen. Dit zijn de zogenaamde kwaliteitscriteria welke duidelijk dienen te omschrijven wat het product moet doen en vaak ook wat het product liever niet moet doen. Dit kunnen voor verschillende producten zeer uiteenlopende eisen zijn, bijvoorbeeld: mechanische sterkte (bromfietshelm), ergonomische eisen (overzichtelijkheid van de regelkamer van een elektriciteits centrale) en eisen m.b.t. de veiligheid. Bovendien kunnen de kwalitietseisen van een product nog verschillen per toepassing en per gebruiksomgeving. Het geheel van eisen bepaalt de kwaliteit van een produkt.

Bedrijfszekerheid is sterk verwant met kwaliteit, zie figuur 1. Simpel gezegd is bedrijfszekerheidstechniek het handhaven van de kwaliteit gedurende een langere periode. Hoe langer de kwaliteit gehandhaafd kan worden hoe beter de bedrijfszekerheid.

Drijfveren voor een bedrijfszekerheidsprogramma

Hoewel het handhaven van de kwaliteit op lange termijn voor sommige producten niet zo interessant is (bijvoorbeeld producten die na kortstondig gebruik weggegooid worden zoals plastic koffiebekertjes) zijn er voor veel andere produkten gegronde redenen aan te wijzen om een bedrijfszekerheidsprogramma op te starten. Een aantal van deze redenen zullen we hier bespreken.

Veiligheid: Wanneer een product bij falen gevaar kan opleveren voor de gebruiker of zijn omgeving kan de fabrikant vaak forse schadeclaims tegemoet zien indien hij niet afdoende aan kan tonen dat hij zijn uiterste best heeft gedaan om het risico zo laag mogelijk te houden door, ofwel het terugdringen van de faalkans ofwel het toepassen van een fail-safe concept. Als voorbeeld kunnen we noemen het verhaal van de schroevendraaier fabrikant die de slijtvastheid wilde verhogen door het gebruikte staal te harden. Het staal wordt dan inderdaad slijtvaster maar ook bros. De fabrikant had echter geen rekening gehouden met het feit dat het algemeen aanvaard is dat schroevendraaiers ook voor andere doeleinden gebruikt worden bijvoorbeeld als breekijzer. Toen een van de gebruikers door een metaalsplinter aan het oog verwond werd is de fabrikant aansprakelijk gesteld. De rechter heeft de eis toegewezen omdat de fabrikant had kunnen voorzien dat ook zijn schroevendraaiers op deze wijze gebruikt zouden worden.

Productieverlies: Het uitvallen van produktie-apparatuur kan tot produktieverlies leiden. Wanneer de producten gemaakt worden op een groot aantal gelijksoortige machines dan kan de uitval van een machine waarschijnlijk wel opgevangen worden zonder noemenswaardig produktieverlies. Echter, in het geval dat het produkt door een enkelvoudig uitgevoerd complex apparaat vervaardigd wordt zoals bijvoorbeeld de kraakinstallatie van een olieraffinaderij dan kan uitval tot een fors produktieverlies leiden. Dit kan nog verergerd worden omdat het opstarten van deze complexe installaties na reparatie vaak nog meerdere dagen inbeslag neemt.

Onderhoudskosten: Ook de beheersing van onderhoudskosten kan een goede reden zijn om een hoge bedrijfszekerheid na te streven. Hoe minder machines een wasmachinefabrikant binnen de garantieperiode ter reparatie krijgt aangeboden hoe minder deze hoeft te investeren in onderhoudsafdeling, reserveonderdelen en transportkosten. Dit zijn tenslotte allemaal kapitaalinvesteringen die geen winst opleveren.

Concurrentie: Wanneer een fabrikant een product op de markt brengt met een hogere bedrijfszekerheid dan dat van zijn concurrent dan biedt dit hem de gelegenheid om zijn marktaandeel uit te breiden ten koste van de concurrentie. Anderzijds, wanneer het product aanzienlijk slechter is dan kan dit een aanzienlijke vermindering van de afzet tot gevolg hebben, zeker wanneer veiligheidsaspecten in het geding zijn. Zo zag een autofabrikant zijn aandeel op de Amerikaanse markt vrijwel totaal instorten toen bleek dat de automatische versnellingsbakken van zijn auto's soms spontaan inschakelden waarbij de stilstaande auto plotseling optrok. Dit heeft tot enkele ernstige ongevallen geleid. Het heeft de fabrikant veel tijd, geld en moeite gekost om niet alleen de foutoorzaak op te sporen en te herstellen maar ook om het publiek weer van de kwaliteit van zijn product te overtuigen.

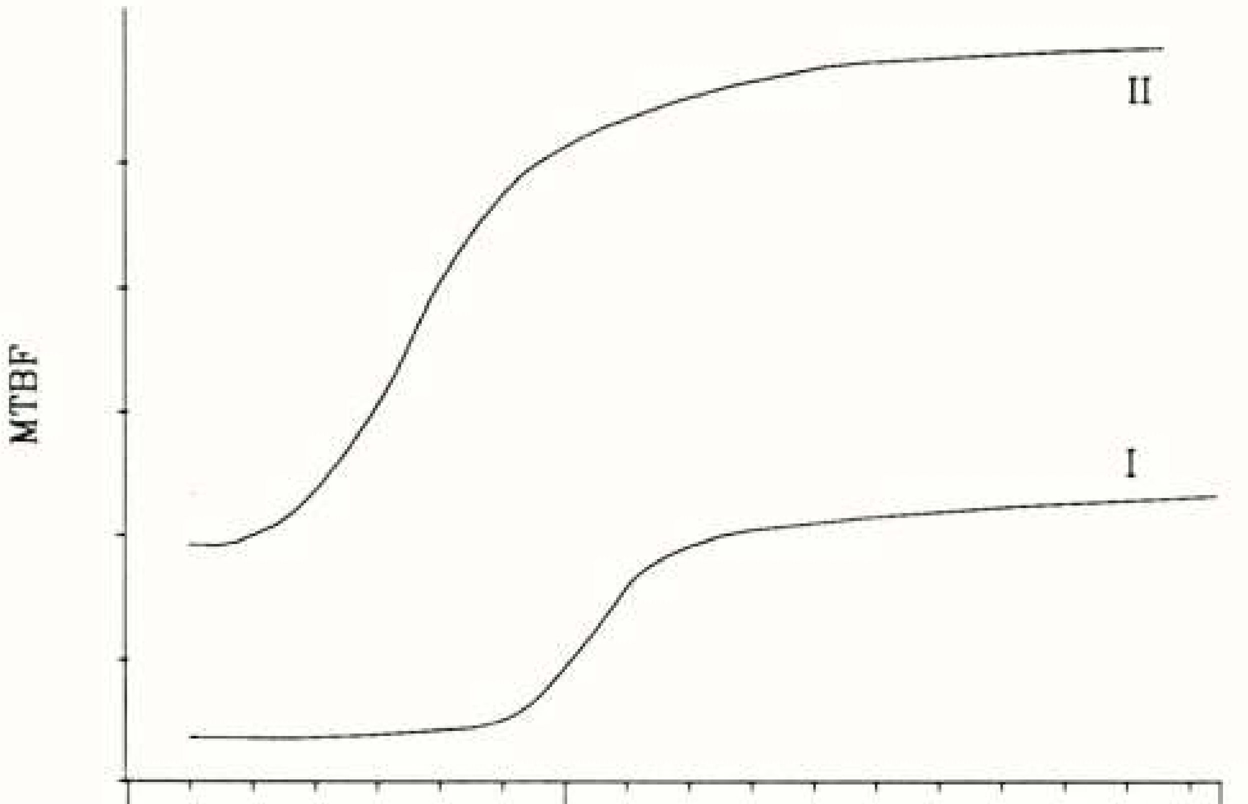


Fig. 2: Twee typische bedrijfszekerheidsgroeicurven voor een produkt. Het is duidelijk dat het bedrijfszekerheidsprogramma zo vroeg mogelijk in de ontwikkelingsfase geinitieerd moet worden teneinde een maximaal rendement te behalen.

Opgelegde eisen: Steeds vaker leggen overheden bedrijfszekerheidseisen (veiligheidseisen) op tafel voor installaties en produkten die schade kunnen veroorzaken aan personen of omgeving. We moeten hierbij niet alleen denken aan grote installaties waarbij grote concentraties gevaarlijke stoffen aanwijzig zijn maar ook aan simpele huishoudelijke apparaten. Wanneer een fabrikant met zo'n eis geconfronteerd wordt dan zal hij deze vaak weer vertalen in bedrijfszekerheidseisen voor onderdelen en subsystemen van toeleveranciers. Een eis kan zo dus tot nieuwe eisen leiden. Ook fabrikanten van consumentenartikelen worden zich meer en meer bewust van het belang van een bedrijfszeker product en zullen dan ook eisen stellen aan de bedrijfszekerheid van toegeleverde onderdelen. De kans dat een willekeurige fabrikant voor de keus gesteld wordt om of aan de opgelegde eisen te voldoen of zichzelf buiten de markt te plaatsen neemt dus steeds meer toe.

<u>Opstarten v/e bedrijfszekerheidsprogramma</u>

Figuur 2 toont 2 typische bedrijfszekerheidsgroeicurven die aangeven hoe de product bedrijfszekerheid toeneemt met naarmate het bedrijfszekerheidsprogramma vordert. Deze groeicurven hebben vrijwel altijd een S-vorm Uit de figuur blijkt ook de invloed van het tijdstip waarop het bedrijfszekerheidsprogramma wordt opgestart. In het eerste geval (curve I) wordt het programma pas gestart wanneer het produkt vrijwel geheel uitontwikkeld is. Het zal duidelijk zijn dat het vergevorderde ontwerp hier beperkende randvoorwaarden oplegt aan de te bereiken bedrijfszekerheid omdat elke modificatie ingrijpende (en dus dure) maatregelen vereist. Bovendien heeft het produkt bij de marktintroduktie een lage bedrijfszekerheid terwijl het produkt juist dan vaak sterk in de belangstelling staat. Wanneer het programma echter direct in het prille ontwerp stadium ingevoerd wordtdan kunnen de verschillende ontwerpalternatieven ook beoordeeld worden op hun consequenties voor de bedrijfszekerheid zodat een meer verantwoorde ontwerpkeuze mogelijk is. De initiele bedrijfszekerheid kan dan al aanzienlijk hoger liggen (curve II) terwijl deze bovendien verder door kan groeien zonder al te grote inspanningen.

Een bedrijfszekerheidsprogramma kost natuurlijk ook geld, maar betaald zich meestal meervoudig terug. Wanneer een bedrijfszekerheidsprogramma tot resultaat heeft dat de MTBF met een factor 20 toeneemt dan zal de onderhoudsfrequentie met eenzelfde factor afnemen wat een besparing van 95% op onderhoudskosten mogelijk maakt. Vanzelfsprekend dient men zich voortdurend af te vragen tot hoe ver de bedrijfszekerheid opgevoerd moet worden teneinde een optimaal resultaat te bereiken. Enerzijds zullen de aanschafkosten Ka van een systeem toenemen met de bedrijfszekerheid, zie figuur 3. Deze kostenstijging ontstaat doordat de ontwikkelingskosten hoger zijn tengevolge van noodzakelijke levensduurbeproevingen, bedrijfszekerheidssimulaties en ontwerpherzieningen. Ook de produktiekosten zullen stijgen als gevolg van duurdere 'high reliability' compo-

nenten, inspectieprocedures, opleiding van produktiepersoneel en dergelijke. Anderzijds zullen de exploitatiekosten K_e (onderhoud, produktieverlies, schadeclaims e.d.) een afnemende tendens vertonen. De gebruiker-eigenaar van het systeem zal daarbij streven naar zo laag mogelijke totaal kosten K_t, betrokken op de gehele levenscyclus van het systeem: ontwerp, produktie en gebruik. In figuur 3 is uitgezet hoe K_a en K_e voor typisch technische systemen afhangen van de bedrijfszekerheid van het systeem. Duidelijk zal zijn dat bij een bepaalde MTBF de totale kosten betrokken op de gehele levenscyclus van het systeem minimaal zijn. Niet alle systemen bevinden zich op of in de buurt van het minimum in de K_t-curve. Rechts van het optimum heeft men kennelijk geeist dat er zeer weinig of geen onderhoud mag worden gepleegd bijvoorbeeld omdat dit niet mogelijk is (ruimtevaart) of omdat falen zeer ernstige gevolgen heeft (kerncentrale). Het andere geval treft men vaak aan bij consumentenartikelen. Voor produkten in de consumentensfeer let de afnemer vrijwel alleen op de initiele aanschafprijs K_a.

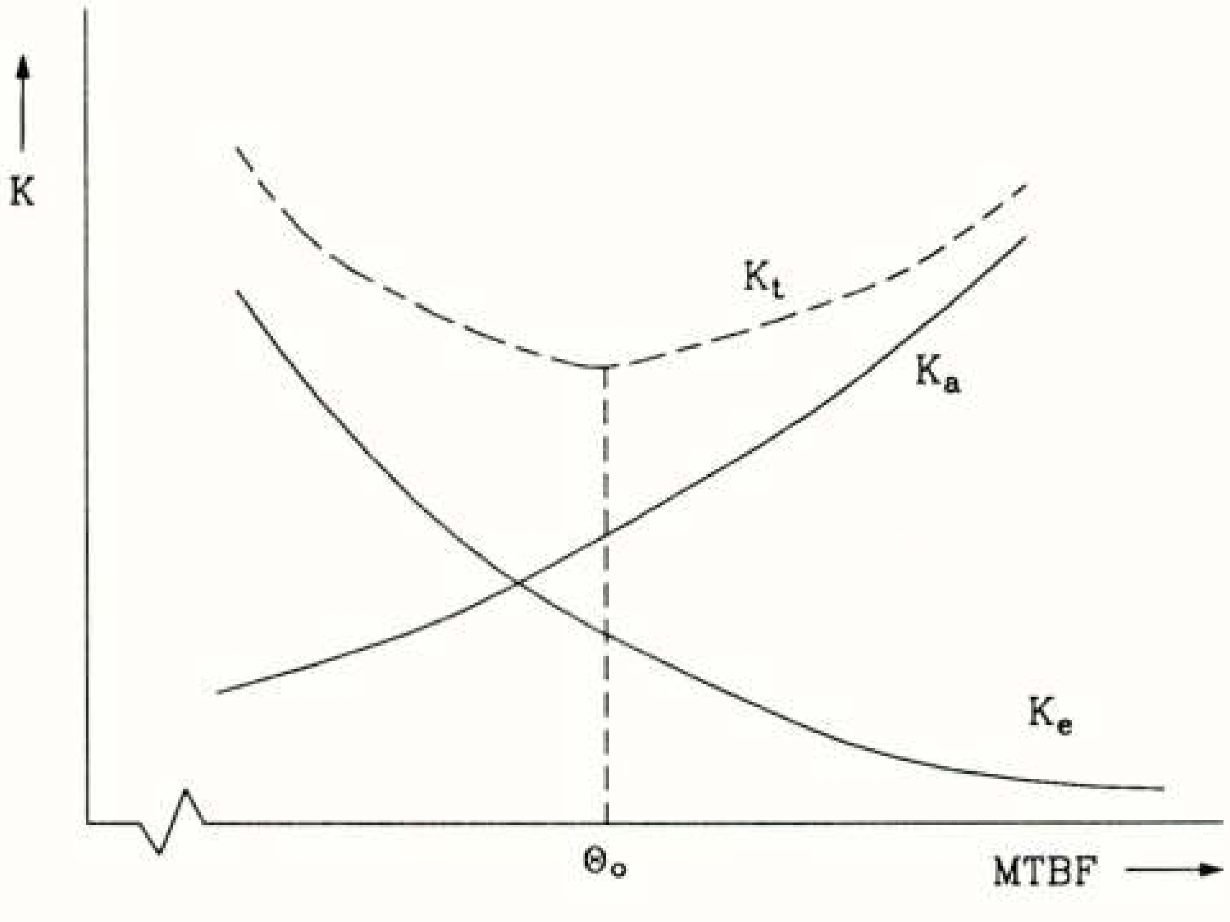


Fig. 3: Aanschafkosten K_a, exploitatiekosten K_e en totale kosten K_t van een fictief technische systeem versus de MTBF van zo'n systeem.

Conclusies

We kunnen nu concluderen dat:

- het streven naar een bedrijfszeker produkt vrijwel altijd een nuttige en rendabele bezigheid is.
- de kosten minimaal zijn wanneer het bedrijfszekerheidsprogramma reeds in het prille ontwerpstadium van het produkt ingevoerd wordt.
- het zinvol is om bij de bedrijfszekerheidsoptimalisatie te streven naar minimale totaal kosten (aanschafkosten + exploitatiekosten) en hier slechts in goed gemotiveerde gevallen van af te wijken.

Literatuur

- Klaassen, van Peppen, Bossche, 'Bedrijfszekerheid theorie en techniek', Delftse Uitgevers Maatschappij, Delft, 1988.

Voordracht gehouden tijdens de 376e werkvergadering.

UIT HET NERG

LEDENMUTATIES

Voorgestelde leden

Ir. E.W. Kolk, Grote Belt 153, 2133 GW HOOFDDORP.
Ir. M.H.A. Paquay, Kwartelstraat 11, 2623 PH DELFT.

Nieuwe leden

Ir. A. van Ardenne, Borghstraat 15, 7841 AL SLEEN.
Ir. E.A. de Boer, Gekerstraat 158, 7523 VT ENSCHEDE.
Prof. dr. ir. R.T. Boute, Pauluslaan 13,
6564 AN H. LANDSTICHTING.
Ir. F.W. Hoeksema, Universiteit Twente, Vakgroep CIS-EL,
Postbus 217, 7500 AE ENSCHEDE.
Dr. phil. H.K. Leonhard, Wijnberg 29, 2716 PE ZOETERMEER.
Ir. H.G. Rave, Vluttersven 19, 5646 HX EINDHOVEN.
Ir. J.H.A. de Rijk, Van Hasseltlaan 309, 2625 HV DELFT.
Ing. G. van Stam, Pruimendijk 184, 2989 AL RIDDERKERK.
Mevr. ir. M.L.G. Thoone, Tiergartenstrasse 22, D6333 BRAUNFELS,
West-Duitsland.
Ing. E.S. Trommel, Ruimtebaan 319, 2728 MP ZOETERMEER.
Ir. M.J.H. van Veen, Schaepmanlaan 3, 4102 BW CULEMBORG.
Ing. A.M. Velders, Philips Nederland B.V., Groep Kabeltelevisie,
Postbus 90050, 5600 PB EINDHOVEN.
J.W. Verhoof, Van Hasseltlaan 250, 2625 HR DELFT.

Nieuwe adressen van leden

A.J. van den Berg, Rivierensingel 747, 5704 KV HELMOND.
D.E. Boekee, Aronskelkweg 6, 2555 GH DEN HAAG
H.L.H.M. Castermans, Egellaan 3, 5581 EG WAALRE.
A.P. Dekker, Kemistintie 7 H 77, 90570 OULU, Finland.
Mevr. ing. E. Harreveld, Van Beethovenlaan 47,
2253 BE VOORSCHOTEN.
Ir. L.L. Kossakowski, Gregoriussingel 52, 5216 GJ DEN BOSCH.
J.P.A. Lamb, Interface Plus, Hoge Morsweg 80, 2332 HM LEIDEN.
W.L. van der Poel, Fagotstraat 18, 2287 BD RIJSWIJK.
G. van Stam, Letlandsestraat 16 B, 3028 EE ROTTERDAM.
Ir. M.J.H. van Veen, Chopinplein 173, 4102 CX CULEMBORG.
Ir. J.A.H. Voorthuizen, Köhlerweg 16, 8024 OBERACHING,
West-Duitsland.
P.P.M. van der Zalm, Nieuwe Parklaan 165, 2807 BS
SCHEVENINGEN.

1990 IEEE International Workshop on Information Theory 10-15 juni 1990, in Veldhoven, Nederland

Van 10 tot en met 15 juni jongstleden is in congrescentrum "Koningshof" in Veldhoven de IEEE Information Theory Workshop gehouden. Er namen 125 mensen deel aan de Workshop, afkomstig uit 19 verschillende landen. De meeste Europese landen (vooral ook Oost-Europa) waren vertegenwoordigd, maar er waren tevens deelnemers uit de U.S.A., Japan, Zuid-Afrika en Israël.

De Workshop omvatte acht sessies, gegroepeerd rond onderwerpen uit de Shannon theorie, magnetische recording, cryptologie, blok- en convolutiecodes, neurale netwerken en broncodering. Elke sessie bevatte gemiddeld zes voordrachten van een half uur. Dat de meeste verhalen zowel inhoudelijk als qua presentatie van hoge kwaliteit waren is niet in de laatste plaats te danken aan de organisatoren van de sessies, die de sprekers uitgenodigd hadden. Extra tijd (50 minuten) was ingeruimd voor A.J. Viterbi (Qualcom, San Diego, U.S.A.), die een overzichtsverhaal hield over de ontwikkeling, in theorie en praktijk, van het door hem ontworpen decodeeralgoritme.

Ook in de rest van de sessies was zowel de theorie als de praktijk vertegenwoordigd. Er waren voordrachten over zuiver theoretische resultaten, bijvoorbeeld over capaciteitsaspecten van het Arbitrarily Varying Channel, maar ook duidelijk op de praktijk gestoelde verhalen zoals dat van T.D. Howell (IBM Almaden Research, U.S.A.) over het "One Gigabit per Square Inch Magnetic Recording System".

Menige voordracht bracht een levendige discussie op gang, die in een enkel geval voortgezet werd tijdens het officiële banket. Zelfs de exotische danseressen, die dit in Indonesische stijl gehouden banket opluisterden, vermochten de aandacht niet af te leiden van het spectaculaire "wire-tap channel" resultaat dat die middag door U.M. Maurer (ETH Zürich, Zwitserland) gepresenteerd was.

Het niet-technische deel van het programma bevatte verder onder meer demonstraties van volksdansen, vendelzwaaien en een draaiorgel (als voorbeeld van oudhollandse digitale recording). De excursies op woensdagmiddag trokken heel wat belangstellenden; en dan met name de "Van Gogh Tour", die de deelnemers kennis liet maken met het land van de "aardappeleters". Ook de simultane schaaksessie, diezelfde avond, was een succes.

Al met al zorgden deze activiteiten ervoor dat de Workshop in een prettige, gemoedelijke sfeer verliep. Bovendien zullen de voordrachten menigeen inspiratie hebben gegeven voor zijn of haar onderzoek. We kunnen dan ook terugkijken op een zeer geslaagde Workshop.

Ineke van Overveld, Technische Universiteit Eindhoven.

Reliability Research on IC Packaging

ir. H.C.J.M. van Gestel

Electronic Instrumentation Laboratory, Department of Electrical Engineering,
Delft University of Technology, Mekelweg 4, PO. Box 5031, 2600 GA, Delft.

The development of IC packages is nowadays of growing interest. This because of the change in electronic systems which nowadays need to be light and compact. These demands cause the types and sizes of the IC packages to change. However these changes affect also the reliability of an IC package. It appears that the type of the IC package used strongly depends on the design constraints of an electronic system. The failure modes of an IC package in return are related to the design of the package. To keep up with the trends in the packaging world the semiconductor industries are putting a lot of effort in optimizing the assembly process for chips. However theoretical models for predicting package failures still remain to be designed. Therefore packaging was chosen amongst other subjects as a research project Universities in order to serve the Dutch industry.

Introduction

In the last decades electronic systems are integrated in silicon at increasing scales. At the moment their is a clear tendency towards integrating electronic systems with still increasing sizes and complexity. The main driving force at this moment to enlarge the electronic circuits on a single chip (also called a "die") is the cost per interconnection, which is illustrated in Figure 1. The curves in this picture clearly show the advantage of silicon integration in decreasing interconnection costs. Hence, it appears that for electronic systems the costs per interconnection are cheapest when as much electronic circuit as possible is placed on a single chip.

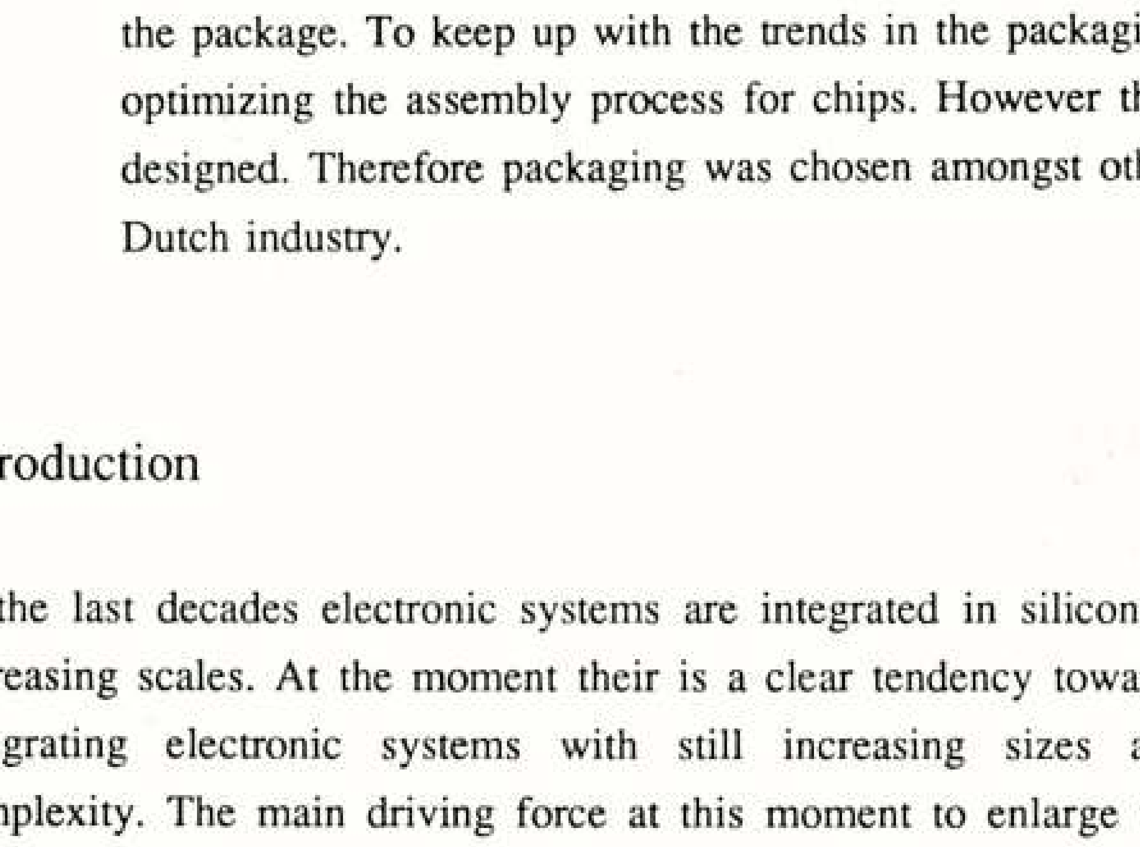


Figure 1 The relative cost of interconnect for Double Sided Rigid (DSR) printed wiring board, Multilayer (MLB) printed wiring board, ceramic hybrids, gate arrays and custom silicon.

The section above gives one reason why IC manufacturers are so willing to integrate large electronic systems on one single chip. However, if the integrated electronic systems are becoming larger the dimensions of the chips (dice) are also increasing. A major result of increasing chip size is the larger becoming chance the chip may not be functional. This may be caused by fabrication failures which occur with a change proportional to the chip surface.

Therefore it becomes necessary for IC manufacturers to make their minimum dimensions for chip processing smaller in order to let the chip sizes decrease and also to improve their production process.

Summarizing both tendencies, previously mentioned, we still observe increasing dimensions for IC's. This implies that a (larger) chip needs more (I/O) connections and hence requires a package that can handle both the (increasing) chip size and the increasing number of pins as well. Therefore new types of IC packages are almost continuously being developed. These new packages require less plastic area compared to both chip area and number of leads. One of the consequences of these new demands on IC packages is that reliability becomes often point of discussion.

Hence, we may observe as driving forces for package related reliability research: increasing chip size and decreasing on-chip device dimensions which results in dice that are more sensitive for their environmental conditions.

General Production of IC Packages

At the moment a lot of different production methods for electronic packages are available. However the cheapest method, which uses epoxy moulding compounds, still possesses about 80% of today's package market. The basic construction of this package is shown in Figure 2. In this picture a number of terms can be found, like die and leadframe, that will be explained in the following sections.

The first step in the assembly flow, (encapsulation process of chips), is to test the chips on the silicon wafers that come from the "fab" and then to saw them into single chips, the so-called dice. Generally the wafer is placed on a foil before sawing, after sawing the foil is stretched so the dice are separated a bit from each other. This facilitates further handling of the dice.

The second step in the assembly flow is to glue the dice to a leadframe. A leadframe is often a copper iron or nickel iron sheet designed for a particular type of plastic package, see Figure 3. From

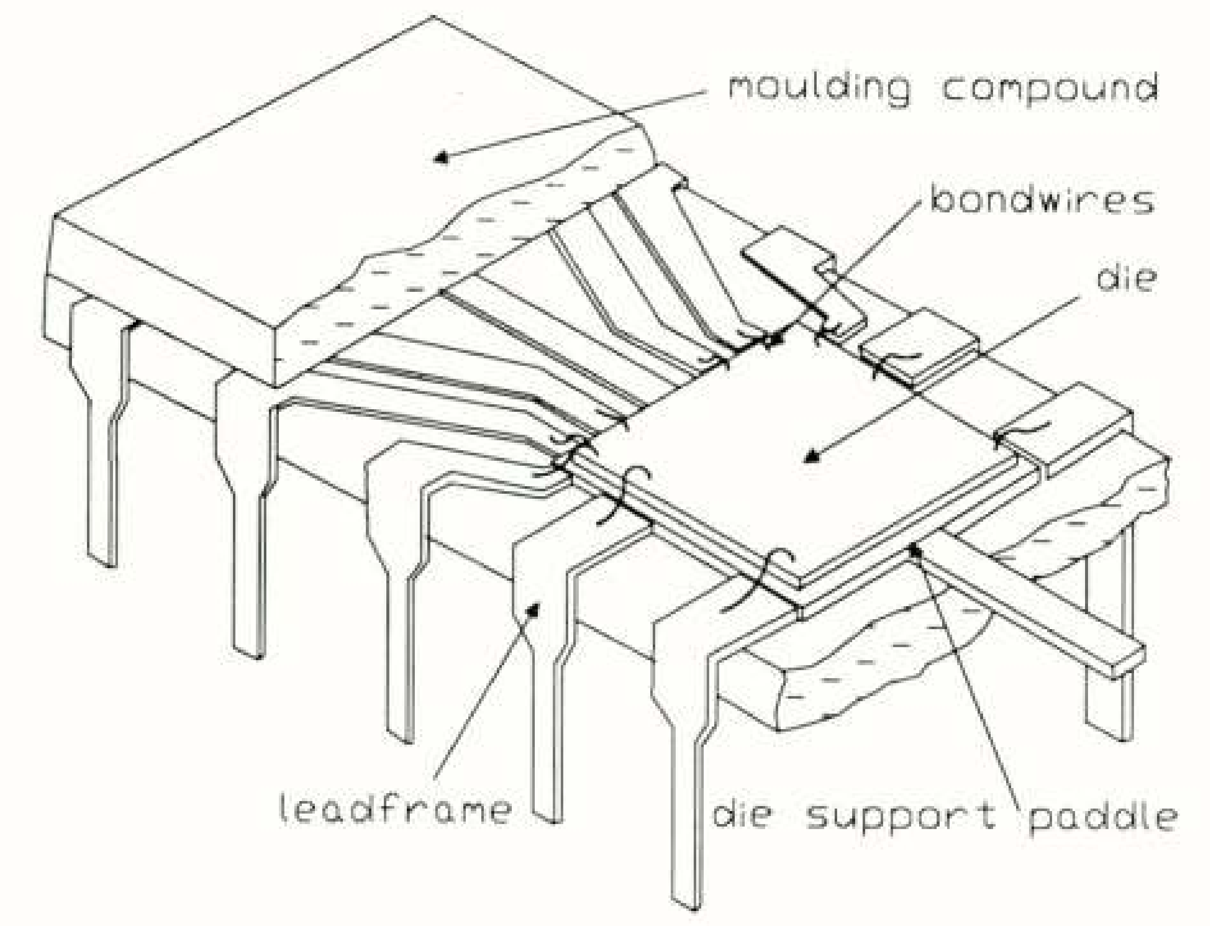


Figure 2 Open view of a dual in line plastic package, often referred to as a DIP or a DIL.

these leadframes the die pad and the pin connections near the die pad, the inner leads, are silver or gold plated, to improve electrical conductivity and adhesion of the die to leadframe, but it also improves (wire) bonding performance at inner leads. In order to glue the dice on the leadframes a silver-filled epoxy is stencilled on the die pads of the leadframes. Then pick and place tools take the dice from the foil and put them on the die pads with the epoxy layer, (polymer adhesive). After this step the leadframes are heated in ovens up to temperatures between 180°C and 275°C for a period ranging between a few minutes up to two and half hours, depending on the kind of adhesive that is being used. This step in the assembly flow is called the die-attach or the die-bond.

After the epoxy is cured the leadframes cool down and are ready for wire bonding. Wire bonding is the most important interconnection method between the silicon dice and its packages. The wirebond is usually performed with gold wire which has a thickness of about 25 to 33 µm, though aluminum wire is being used also. Generally wire bonding is performed by using thermo-sonic and/or thermo-compressive techniques, in order to obtain a so called "ball-wedge" bond, or also called a "ball-stitch" bond.

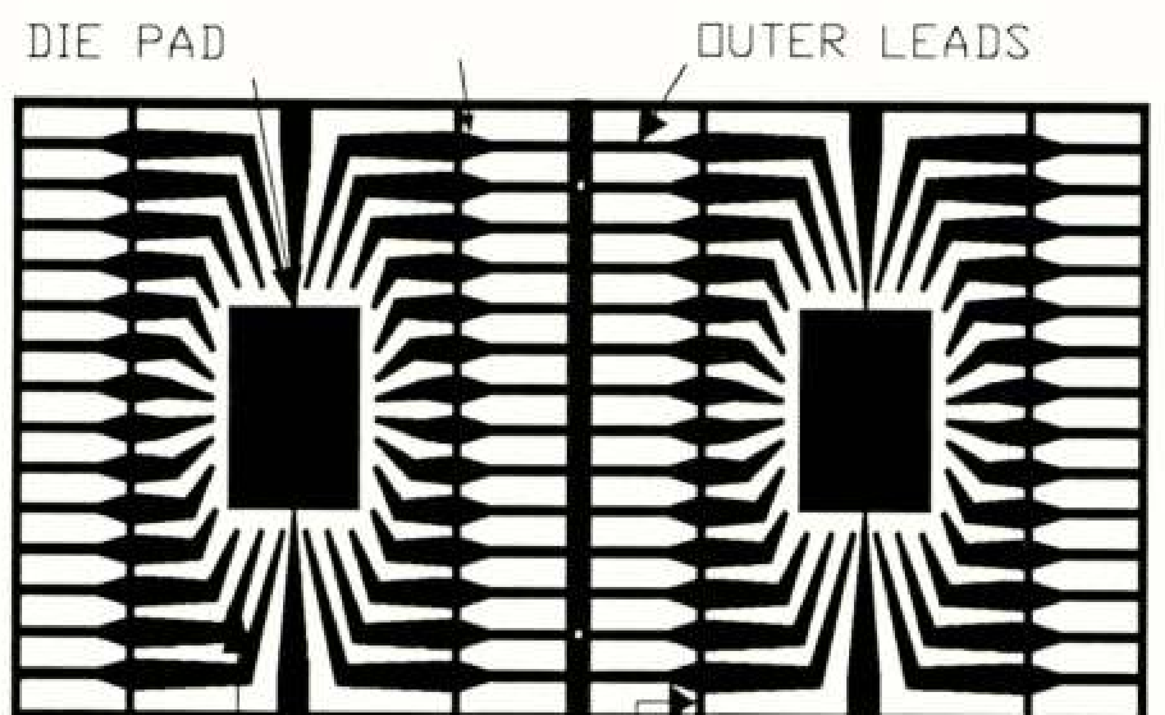


Figure 3 Schematic drawing of a part of a leadframe.

After die-attach and wire bonding the chip/leadframe combinations are transfer moulded with means of an epoxy (moulding compound). Usually thermosetting materials such as epoxies are used as moulding compounds, but these epoxies always contains a large number of ingredients to adjust the epoxies for the needs of IC packages. The thermosetting material is preheated and then enters the pot under pressure and heat. Then by lowering a plunger the encapsulation material is forced to flow through the runners and the gates and then it fils the cavities containing the chip-leadframe combinations, see Figure 4. The IC leadframe often has long, fragile lead fingers, and

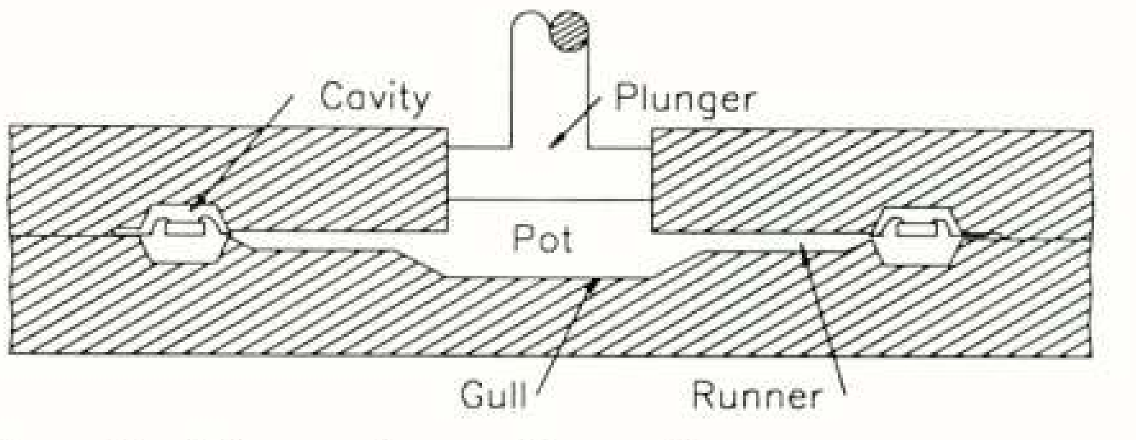


Figure 4 Scheme of a transfer moulder.

the die is interconnected to these thin fragile leads with 25 to 30 µm diameter gold wire. To avoid damaging this fragile structure, the viscosity and velocity of the moulding compound must fall within certain ranges. Some productivity problems that may arise during transfer moulding are: flash on the lead, wire sweep, inner voids or inner lead deformation. Commercial moulding compounds are designed to meet these requirements when moulded at approximately 180°C and at pressures of about 6 MPa. The mould cycle requires from 1 to 5 minutes. To control the velocity of the moulding compound, each device cavity has a gate or restriction to slow the material flow. The fluid mechanics of moulding are relatively complex, because the materials are non-Newtonian. In addition, partial cross linking can occur during the moulding process affecting the material viscosity. Elaborate mould designs have been proposed to compensate for these variables.

After curing the packages are cut loose from the leadframes, the plastic is removed from the connection leads (deflash) the cross lead support and dam-bars are also cut away and the leads are bent down and tin-plated, or tin-dipped.

Trends in Packaging

It appears from Table I that the transfer moulded IC package nowadays represents 85-95% of all semi-conductor output from Japanese companies. The rest is split between ceramic packaging, chip on board and tape automated bonding (TAB) technology. It also appears that new moulded package types are being developed almost continuously. The driving force for new package types depends on the product and its application. For example Gate Array product requirements are high integration, and high pin count packages whereas Dynamic RAM products require high density, and small thin package types. These requirements not only form the driving force to

develop new package types but causes new materials to be developed as well. Also heat dissipation performance and thermal resistance of IC packages is of significant importance for some manufacturers and causes the development of new package materials too.

Nowadays the variety of package styles is split between the traditional insertion type and the new surface mountable packs. From Table I it appears that surface mount packaging represents 20% - 30% of total output at the present time and is increasing year by year.

Table I Japanese IC packaging

Company	Total usage in House	Plastic	Surface Mount
Hitachi		95%	30%
Nec	20%		20%
Mitsubishi		95%	20%
Fujitsu		85%	20%
Oki	10%	90%	50%
Matsushita	40%	95%	20%

Package Design

The purpose of IC packaging is in the first place to protect the die from environmental conditions or circumstances. Secondly the package provides a possibility for making electrical contact with the bond pads on the chip. And a third reason is to provide the chip a larger cooling surface or in other words, to provide a lower thermal resistance from the chip to its environment. One package type however cannot meet all of these requirements at the same time. Therefore a choice of the package type has to be made. Usually this is done by the IC supplier, based on parameters such as die size, I/O count, power dissipation, speed,environmental requirements and costs. Generally we could describe some design constraints for a particular IC for a particular application. Most of the constraints may be considered to satisfy one of the following constraint fields: geometrical, electrical, thermal, mechanical, and cost (to manufacture and support).

Examples of geometrical constraints are e.g. the number of IC connection pads related to the dimensions of the die which in it's turn is related to the number of gates on the chip. It also demands a certain number of I/O leads and hence requires a particular package. This package must then be placed on a printed circuit board (PCB) and requires a certain area and a number of interconnection layers.

Several electrical performance criteria of interest are for example: low ground resistance (minimum power supply voltage drop), short signal leads (minimum self inductance), minimum power supply spiking due to signal lines simultaneously switching, short-paralleled signal runs (minimum mutual inductance and cross talk), short-length signal runs near a ground plane (minimum capacitive loading), and the maximum use of matched impedances to avoid signal reflection. These criteria are of course not all mutually independent. They may be related through simple geometric variables, such as conductor cross section and length, dielectric and dielectric constant of the packaging body.

The thermal field demands a particular thermal conductivity from the various kind of packaging materials. To keep the junction temperature on chip as low as possible the so called "thermal resistance" of a package must be also be as low as possible. The thermal resistance is the ratio of the temperature change necessary to move thermal power across a heat transfer element. In order to obtain the total thermal resistance a number of environmental conditions must be established like: printed circuit (or wiring) board temperature, total power dissipation on the board, conductivity of the PCB, and ideal performance of the isolated package. Conventionally, either junction to ambient or junction to case (the body of the packaged device) thermal resistance (in $^{\circ}C/W$) is being used. These resistances are usually derived by suspending the device in still air or mounting in a particular socket arrangement.

Mechanical constraints are often related to the thermal expansion coefficient from the different packaging materials and the dimensions of chip and package. Especially large dice the choice for materials like leadframe, die attach adhesive and (low stress) mould compound are very important. If one doesn't pay attention to this a package is likely to fail very soon.

And yes finally the most important constraint, certainly for mass production, are the costs. Costs may be separated for clarity in wafer fabrication costs, package costs, heat removal costs and printed circuit interconnection costs. For viewing only the IC packaging, we may observe that the costs of packaging for a single chip may exceed those of the chip's production. This is also one good reason for trying to develop cheaper packages with still a high level of reliability.

Accelerated Testing

If we want to accelerate failure mechanisms of a device, we have to harden one or more of the physical quantities that is/are of influence on the lifetime of the device. Such a physical quantity is called a stress. There are four common stresses used to accelerate device failure mechanisms: temperature and temperature cycling, voltage, current and humidity. Temperature cycling is mainly used to accelerate mechanical failure of the chip or assembled package.

In any study performed under accelerated aging conditions, different failure mechanisms may be accelerated by different amounts for the same applied stress. Suppose a device may fail at normal operating conditions because of two completely different failure mechanisms. Under the applied stress, one of these failure modes may be accelerated much more than the other. Thus in these accelerated aging studies, we would see only one failure mode, and we may successfully eliminate that one mode. However under normal operating conditions, we would have reduced the device failure rate by only a factor 2.

Accelerated aging is a useful tool only if we know the device failure mechanisms and the acceleration of these mechanisms as a function of the applied stress. Adequate studies must be done under normal operating conditions to satisfy ourselves that no failure mechanisms remain that were not accelerated by the applied stress. The ultimate test of device reliability is long hours of operation under expected conditions of normal use.

Temperature Acceleration

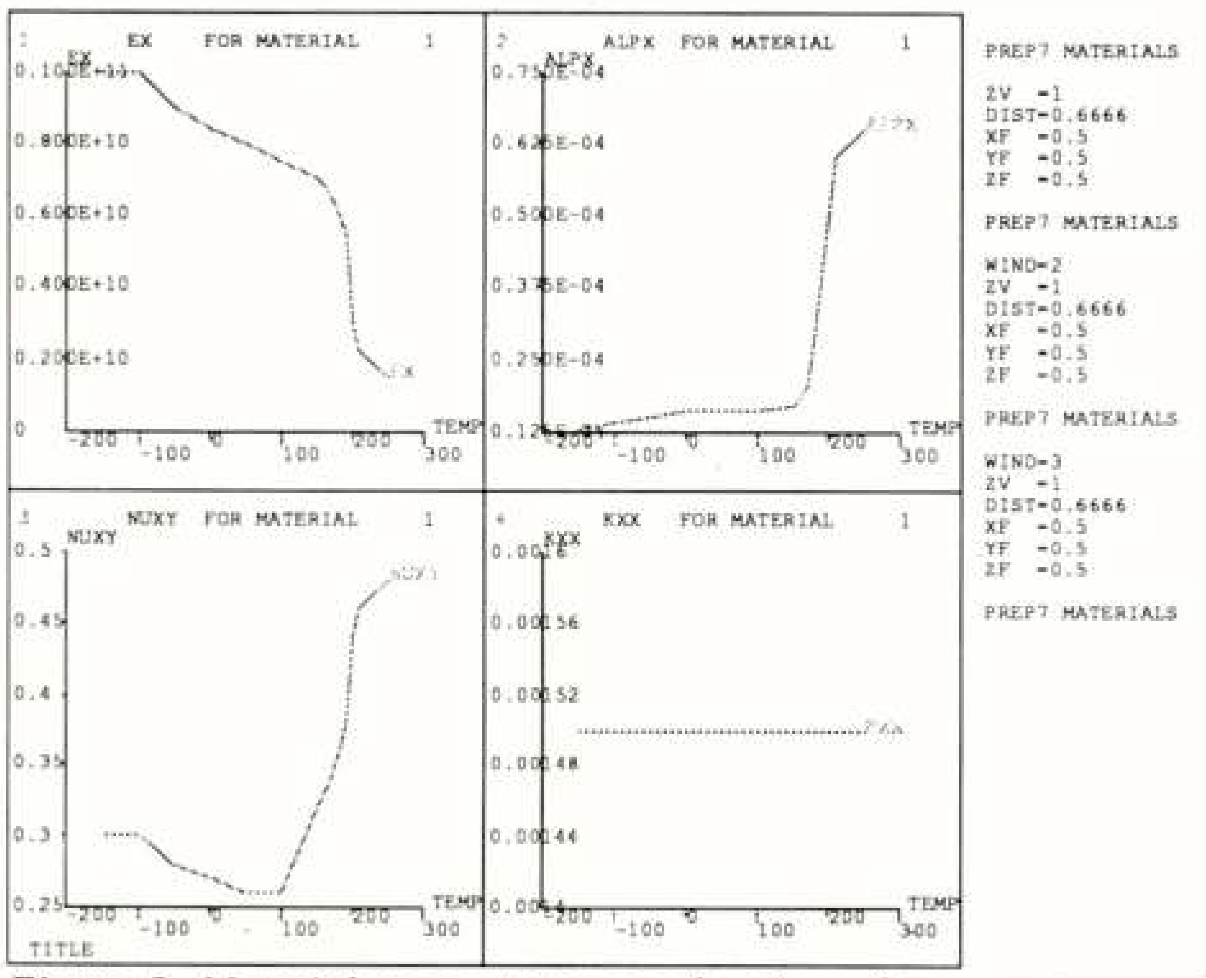


Figure 5 Material parameters as a function of temperature used for finite element calculations.

Many of the mechanisms that cause failure are chemical or physical processes that can be accelerated by temperature. The reaction rate at which these processes proceed is governed, in many cases, by the Arrhenius equation. Therefore the so called burn in of IC packages is often performed at a high temperature (up to 150 C).

Due to the large mismatch in thermal expansion coefficients of the materials used in an IC package large stresses may arise in the package as a result of die adhesive and moulding compound curing. In Figure 5 some material parameters of a moulding compound are plotted as a function of temperature. These parameters can be used e.g to simulate the mechanical stresses that arise in the package as a result of the cure process in the transfer moulder. The results of such an analysis with means of a three dimensional model are illustrated in Figure 6, Figure 7 and Figure 8. Note that these figures represent only one quarter of the real die, die adhesive and die support paddle.

Also a well known stress test is the temperature cycling test (TCT). In such a test the results of material degradation can be studied. These degradation effect may cause several possible package related failures. One of the most important results of TCT is delamination of layers with (normally) initially good adhesion. Some results of e.g. delamination of the moulding compound to the chip surface are bondwire failures, pattern shift (of the interconnect on top of the chip) or top to bottom shorts (in case of 2 interconnection layers). Temperature cycling also causes delamination at the die bond adhesive

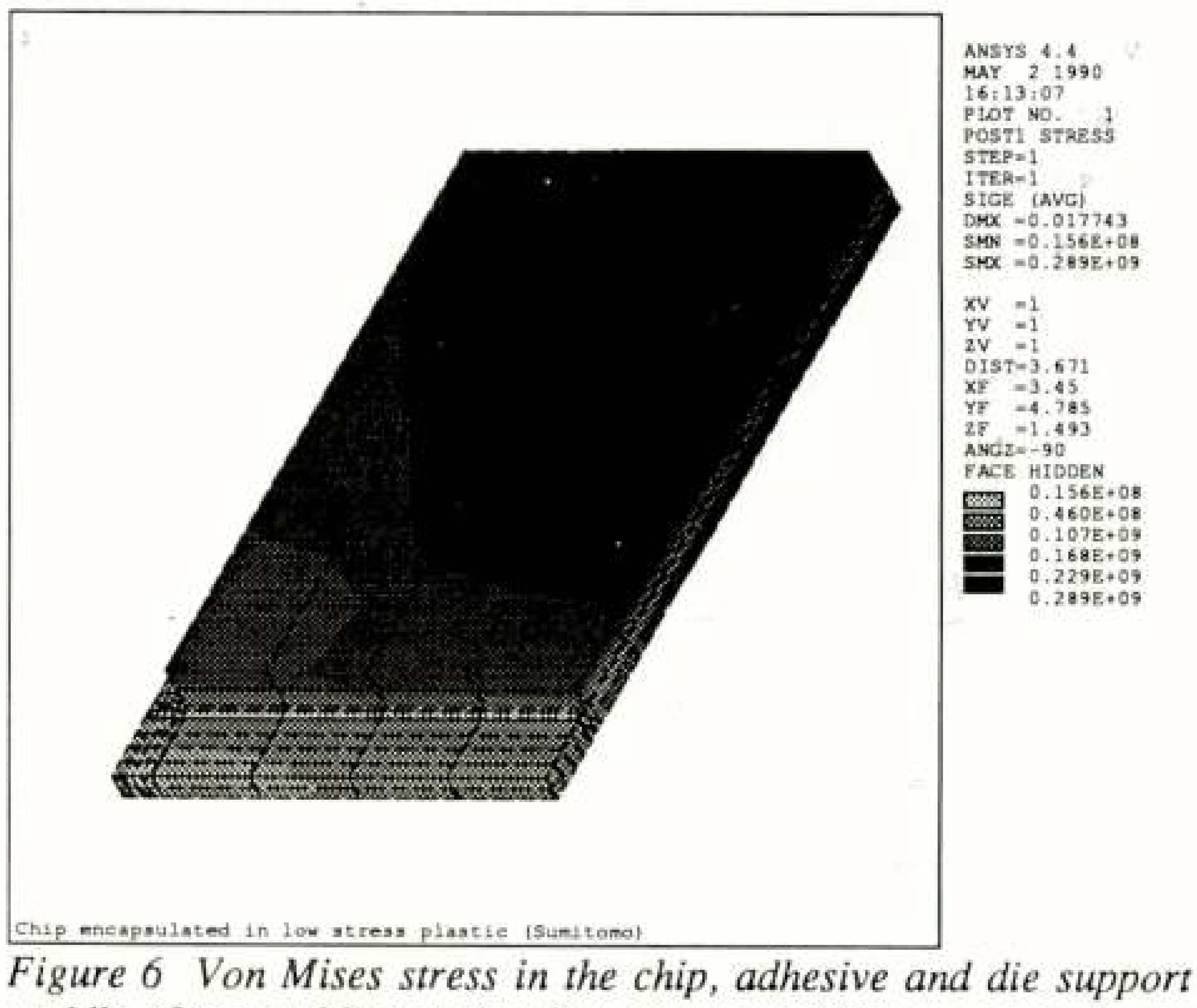


Figure 6 Von Mises stress in the chip, adhesive and die support paddle after moulding with a low stress moulding compound.

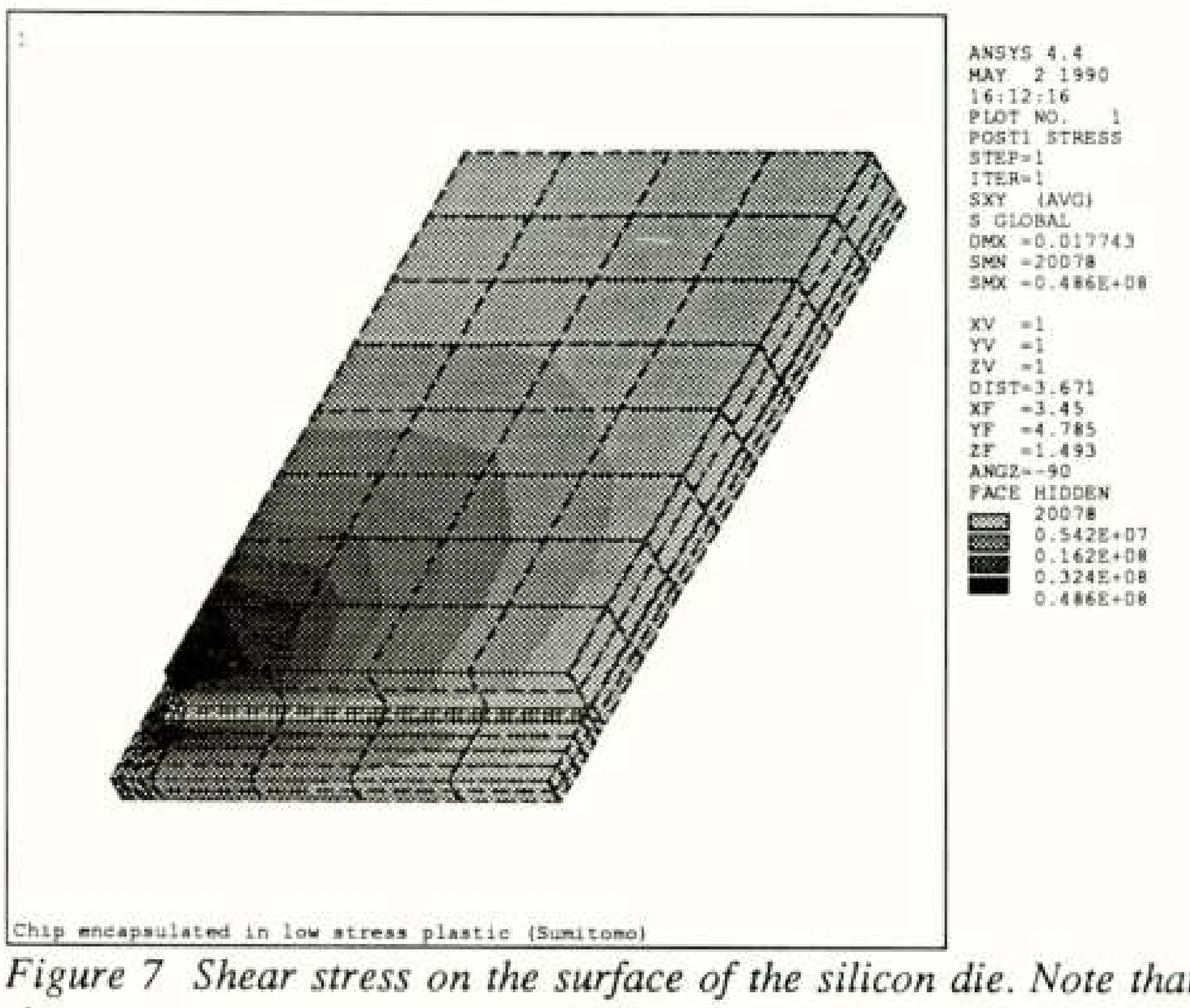


Figure 7 Shear stress on the surface of the silicon die. Note that these stresses are most severe in the "free" corner of the chip.

or at the interface between die support paddle and moulding compound. It has been shown that both delamination mechanisms previously mentioned may cause cracks that originate from the corners of the die paddle and propagate to the surface of the package. There are two main reasons that a plastic crack arises. One is the large mechanical stress that exists in the plastic near the corners and edges of the die pad and the other is the absorption of moisture that causes a crack when the package is heated ("pop corn").

Another purpose of cycling the packaged devices over an extended temperature range is eliminating potential infant mortality failures (burn-in) from the devices shipped to customers. Failures such as weak wire or chip bonds and partially cracked chips or packages may be eliminated by this test.

Voltage and Current Acceleration

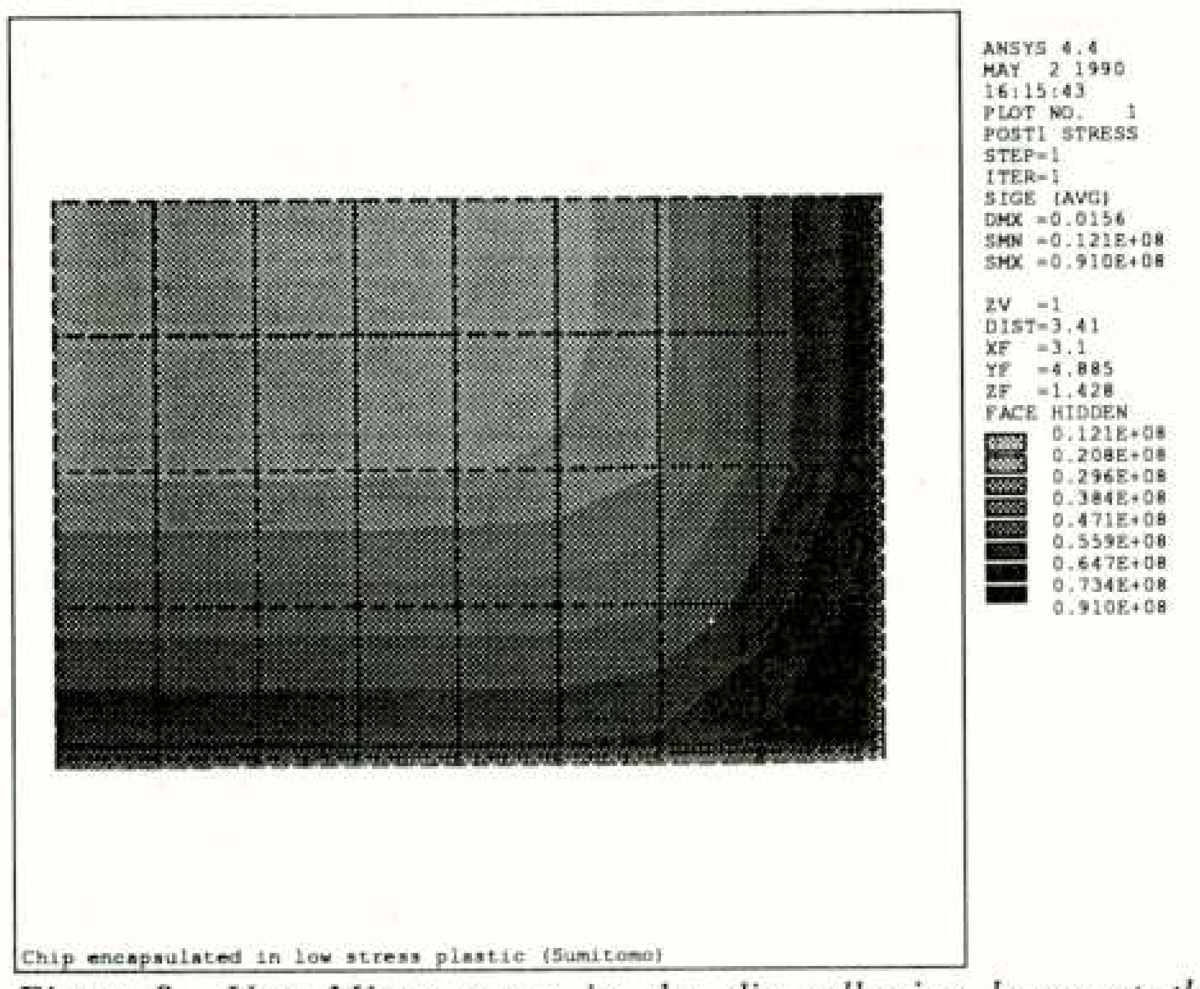


Figure 8 Von Mises stress in the die adhesive layer at the interface of the silicon die.

Voltage and current are effective accelerating stresses for many of the common failure mechanisms observed in ICs. Voltage (or in some case electric field) causes acceleration of failure caused by dielectric breakdown, interface charge accumulation, charge injection and corrosion. With most ICs, varying the applied voltage over a very large range is not possible.

Humidity-Temperature acceleration

Usually, state-of-the-art ICs are packaged in ceramic hermetic packages. These ceramic packages show a higher level of reliability especially if we are dealing with moisture penetration. However if the devices mature, competitive pressure to reduce costs results in the use of lower cost, nonhermetic plastic packages. The presence of water vapour quickly permeates plastic packaging material. In the first step of the permeation process, water vapour transports contaminants from the surface of the package through the plastic and leaches impurities from the plastic packaging material itself. The surface of the chip is very quickly exposed to the water vapour and various contaminants. At this point leakage currents already may increase inside the package.

The second step is the diffusion of the contaminated water vapour through the passivation layer of the chip, usually a relative slow process. However, if the passivation layer contains defects or cracks the water vapour penetrates through the passivation layer much faster. In these studies the penetration of water vapour through the passivation layer determines the reaction rate of the failure mechanisms.

Once the water vapour reaches the metallization level of the chip, electrochemical corrosion can occur. The ions needed for this corrosion process arise from two sources. As described above, ions can diffuse as a contaminant through the passivation layer along with the water vapour. However, if the intermediate dielectric of the chip is a phosphorus-doped glass, the water vapour can leach phosphorus from the intermediate dielectric. In either case, electro-chemical corrosion is a very rapid process, which results in metallization failure. This type of failure mechanism can be accelerated by increasing the partial pressure of water vapour in the environment. Increasing the ambient temperature also increases the rate limiting diffusion of the water vapour through the passivation layer.

Another failure mode which is related to moisture absorption is the so called "pop corn" effect. The effect is sometimes observed when mounting surface mounted devices (SMD) on a PCB. This basic crack mechanism starts with moisture absorption prior to assembly. This causes increased stress in the moulding compound during the soldering operation because of increased vapour pressure. Local swelling is often the result, due to delamination between the epoxy and the underside of the leadframe or above the chip. Crack propagation then starts at positions of maximum stress which are at the bottom and top edges of the chip-leadframe assembly. From experimental reliability tests it appears that pop corn very much depends on the die size and the thickness of the package. Several trends like smaller and thinner SMD packages, relatively larger silicon chip sizes and increasing use of infra-red reflow soldering causes the seriousness of this problem to increase. Hence, to prevent package cracking during the solder reflow the user is therefore often advised to "bake" his packages first before mounting them on the PCB. This preheating step (+100 C) causes enough moisture to evaporate to reduce its concentration in the moulding compound to a save value. Basic material factors which are important to prevent pop corn are the moisture absorption rate of the epoxy, the epoxy to leadframe or epoxy to chip adhesion and the choice of resin which minimizes stresses in the moulding arising from mould shrinkage and heat exposure.

α-Particle-Induced Soft Errors

Radioactive elements, such as uranium or thorium, are naturally occurring impurities in IC packaging materials. The a particles emitted by these materials can cause soft errors in the ICs. The term "soft error" refers to a random failure not related to a physically defective device. The penetration of an a particle into the silicon causes the generation of an electron-hole plasma along the path of the particle. The generated carriers can cause the loss of information stored in the memory cells of a dynamic memory or stored in the depletion region of the drains of the devices making up circuits such as dynamic shift registers or other logic units. The adsorption of a 4 MeV a particle can generate 10 electron-hole pairs, which is equal to or greater than the charge stored in a dynamic memory cell. To illustrate the seriousness of the problem, a typical soft error rate in a memory system containing 1000 16K memory devices may be on the order of one soft error per 1000 h. The incidence of soft errors can be reduced by surrounding or coating the IC chip with a material having a very low density of radioactive contamination. For example, a particles with an energy up to 8 MeV are completely absorbed in 50 μm of silicone rubber, and this material does not emit any significant amount of a particles.

Test chips for evaluating package related failure mechanisms.

At the electronic instrumentation laboratory we are working on test structures for measuring several aspects of package related failure mechanisms. Currently we are developing sensors measuring mechanical effects such as pattern shift and mechanical stress. For the pattern shift sensors we have both a magnetic approach based on the Hall-effect and a capacitive approach. The Hall sensors are placed in a matrix providing information of possible pattern shift in two directions with a measure range up to 10 μm and a resolution of less than 0.5 μm. For the mechanical stresses we are working on a stress test chip capable of measuring not only the in-plane mechanical stress but also the stress perpendicular to the chip surface. The stress sensor is based on the piezoresistive effect. Future work will also include moisture and corrosion sensitive devices, dielectric sensitive transducers for (post) cure measurements, switching logic for electrical characterization and temperature test structures for thermal characterisation.

Improvements Towards Packaging Materials

At the moment a lot of effort is being put in packaging materials improvements. Progressions may be observed in the moulding compounds, die bond adhesives and coatings.

Improvements towards the moulding compound for IC packaging have directed attention to ionic purity levels, the reduction of internal stress created by the encapsulant, minimizing water absorption properties and improving adhesion strength between surfaces. An approach of reducing stress within the mould material is e.g. to reduce the elasticity modulus and the thermal expansion coefficient without reducing the Glass Transition temperature. This is being achieved by concentrating on epoxies modified by the addition of Silicone polymers, and then additionally enhanced by adding new synthetic fillers. These new filler particles used in low stress formulations are predominantly spherical and smaller in shape compared to the normal angular shape and random size, thus helping to reduce particle damage to the chip surface.

To improve the adhesion strength between the lead frame paddle and the encapsulation resin, some companies have also developed polyimide primers. Such materials are applied to the back surface of the paddle during assembly in order to reduce delamination problems. Note that most of the new low stress moulding materials are not available on the open market.

Another improvement to protect the die for internal stresses that may damage the chip or even may cause the die to crack is the use of low stress glue. Since the trend for dice is increasing size, mechanical stresses due to the die bonding may cause cracks in the dice more easily. Consequently it may be necessary to use low stress die attach glue for very large dice which helps minimizing the stress in the die also after mould curing. Another approach for decreasing the influence of the mould shrinkage and damage from filler particles is the use of chip (silicone) or wafer (polyimide) coatings.

Summary

The trend in modern VLSI technology has the tendency to enlarge the number of problems of maintaining high yield and reliability. Progress in VLSI technology is achieved at the cost of finer dimensions, larger chip area, more complex processes, and the introduction of new and more complex material systems. All of these inevitably lead to initially higher defect density. Identifying, characterizing, and ultimately eliminating the causes of failure seems necessary part for the development of the packaging trends nowadays observed.

The combined efforts of technology development along with yield and failure mode analysis will provide the technology needed to produce VLSI packages with a predictable life time and a high reliability level.

Voordracht gehouden tijdens de 376e werkvergadering.

HIGH SPEED DIRECT DETECTION TRANSMISSION TECHNIQUES
FOR THE BROADBAND OPTICAL LOCAL LOOP

A.M.J. Koonen
APT Nederland B.V.

Optical fibre transmission systems using direct detection techniques are able to deliver broadband services at multigigabit per second data speeds in local access networks. A review is given of network topologies, system components and the limits they put on the system performance. Results obtained in various laboratories up to now show that direct detection local loop systems with speeds exceeding 10 Gbit/s can offer economic network solutions by using potentially low-cost (opto-)electronic components.

1. Introduction

Worldwide the demand for transportation of higher bitrates is growing. At business customers, local area networks are carrying 10 Mbit/s (Ethernet) and 100 Mbit/s (FDDI); in the future 1.2 Gbit/s (Vectornet) is foreseen. At residential customers, narrowband ISDN utilizes 144 kbit/s. Digital TV can be transported with 70 Mbit/s per videochannel, and it is expected that the promising HDTV encoded in digital format will need at least 150 Mbit/s per channel. For reasons of privacy, picture-in-picture, etc. it is often assumed that video services should be offered non-switched to every residential subscriber. Thus the distribution of e.g. 64 HDTV channels requires a bitrate on the subscriber access line of about 10 Gbit/s ! Clearly, non-switched video distribution is the prime driving force for very high bitrates in the local loop. This paper will deal with network topologies and multigigabit/s direct detection transmission techniques suitable for the distribution of very high bitrate services in the local loop network.

The current RACE program foresees introduction of the Integrated Broadband Communication Network all over the European community, starting in 1995. According to a recent study in Japan, 46% of the Japanese households will have an HDTV receiver by the year 2000.

2. Network topologies

Up to now optical fibre systems were mainly applied in the trunk part of the network, connecting telephone exchanges with point-to-point links. Emerging applications in the local loop use a variety of topologies : tree-and-branch, full star, bus, ring, and mixtures of these.

Full star networks, where individual fibres run from the local exchange to every subscriber, offer maximum privacy and maximum service upgradability, but are also the most costly to install.

Bus and ring networks are cheaper to install, but upgradability and privacy are less favourable.

The most suited network topology for distributive services is the tree-and-branch one, where a single fibre runs from the local exchange to a distant terminal, from which individual fibres run to every subscriber. This topology yields minimum installation costs. Interactive services can be added using multiple access techniques such as time division multiple access (TDMA). A distinct timeslot is allocated to every subscriber, which cannot be grabbed by other subscribers. In this way a logical star configuration is implemented on a physical tree-and-branch configuration. Injection and extraction of the interactive services into the network is done with wavelength division multiplexing devices. In the RACE program these devices have been defined as optical line inlet (OLI) and outlet (OLO). It is foreseen that the interactive services will be carried on a wavelength around 1.3 μm, and the distributive services around 1.55 μm. Figure 1 shows a schematic network layout. An optical power splitter (OPS) is used in the distant terminal (DT) to distribute the signal from the feeder fibre over the drop fibres to every

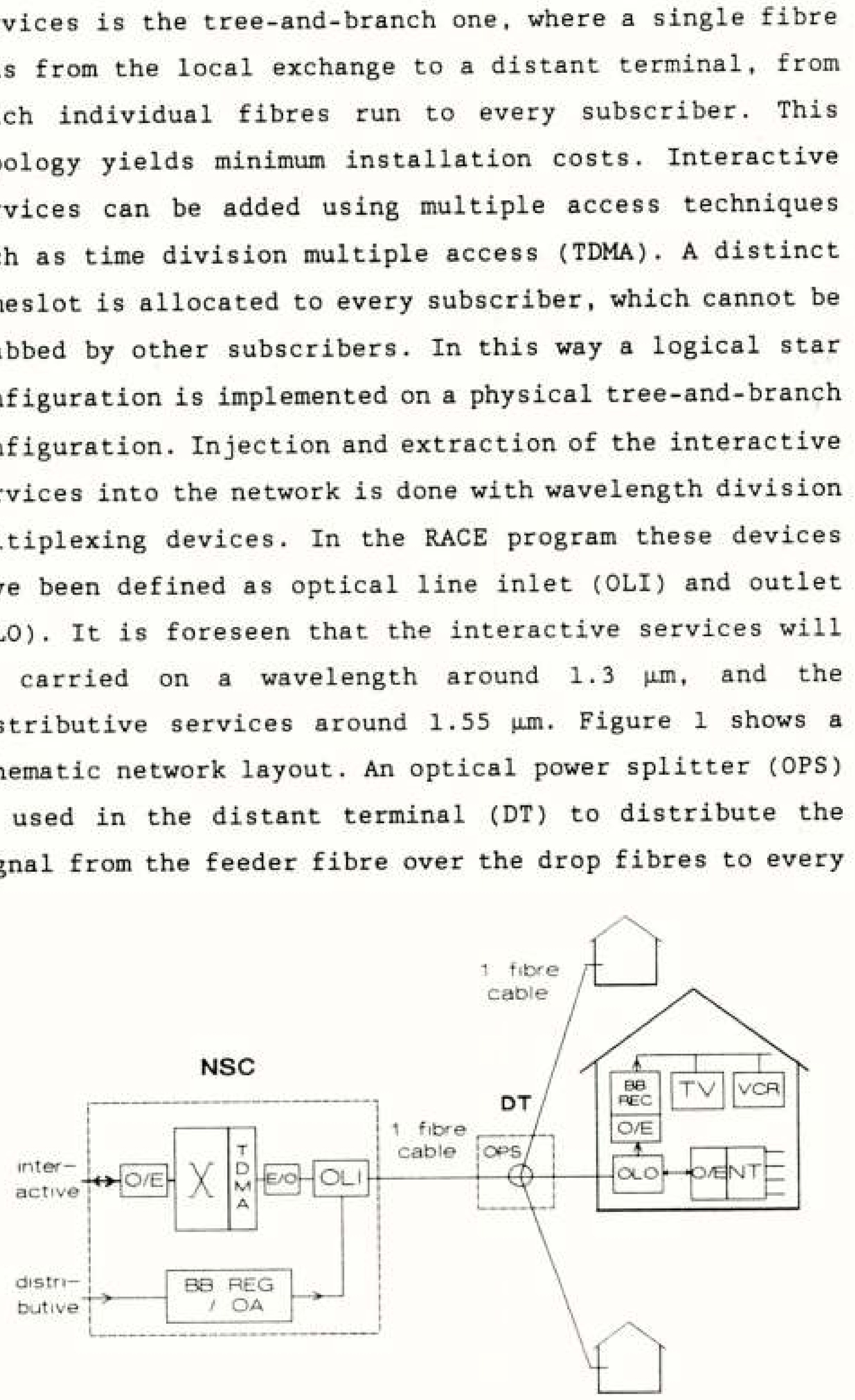


Fig. 1. Tree-and-branch FTTH network integrating interactive and distributive services

household. Thus the network itself is totally passive, which will minimize its maintenance. To reduce the costs per subscriber as much as possible, the equipment in the network service centre (NSC) should be shared by as many subscribers as possible. The splitting factor of the OPS and thus the power budget available between NSC and individual household should be maximized.

3. System components

Figure 2 shows the scheme of a basic high speed optical fibre link, illustrating the line terminal functions. At the transmitter side, N input data streams are multiplexed into the multigigabit/s line signal. A laser driver

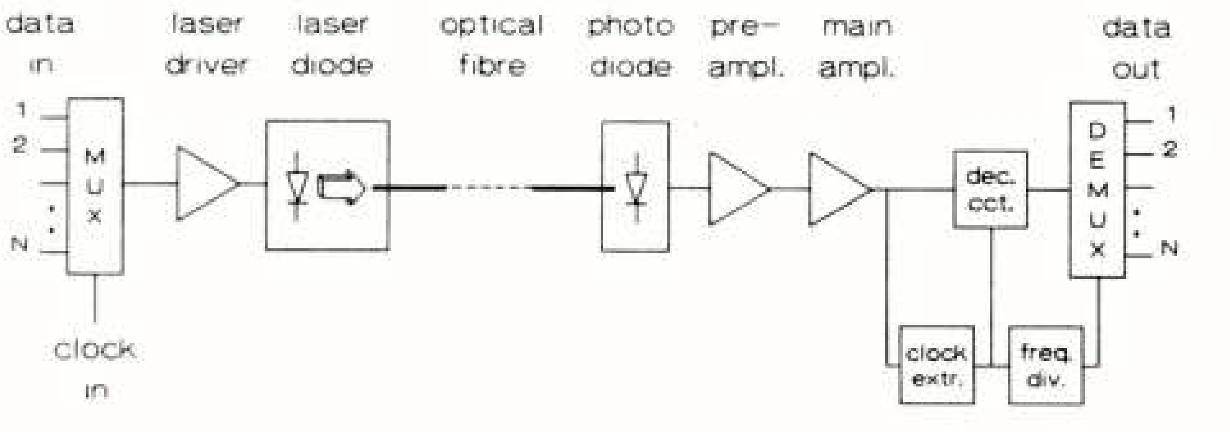


Fig. 2. Basic optical line terminal circuitry

circuit provides the laser diode with the required modulation current. To prevent reflections into the laser, it should preferably have an internal optical isolator. At the receiver side, the signal current out of the photodiode is fed via a low-noise preamplifier to a main amplifier with automatic gain control (AGC) or limiting action. From the resulting constant-amplitude signal the clock extractor regenerates the clock frequency signal, with the aid of which the decision circuit reconstructs the original data. The demultiplexer decomposes the data into the N low bitrate data streams. Its clock signal is derived from the high speed clock signal by a frequency divider.

Table 1. Basic silicon ICs for digital optical fibre transmission systems (cf. [1])

Circuit	Experimental results		Simulation predictions
	Adv. techn.	2 μm techn.	1 μm techn.
MUX	11.4 Gbit/s (2:1)	6 Gbit/s (4:1)	24 Gbit/s (2:1)
Laser driver	–	3 Gbit/s (30 mA)	11 Gbit/s (30 mA)
Preamplifier	3.6 GHz (Zt= 468 Ω)	–	6 GHz, 8 Gbit/s (Zt= 566 Ω)
Main ampl.			
∘ AGC	–	3 Gbit/s, 2.5 Ghz (max. 40 dB)	–
∘ Limit.	–	4.4 Gbit/s (max. 54 dB)	13 Gbit/s (max. 40 dB)
Decision cct.	8.1 Gbit/s	3.8 Gbit/s	15 Gbit/s
DEMUX	11.2 Gbit/s (1:2)	6 Gbit/s (1:2)	24 Gbit/s (1:2)
Static freq. divider	15 GHz (16:1)	3.5 GHz (2:1)	14.6 GHz (2:1)

The electronic circuits should be integrated as much as possible to yield maximum system economy. Gallium-arsenide technology has proven its suitability for high speed integrated circuits, but is still rather expensive. Silicon technology is much more mature, and most suited for lower speed ICs. However, recent progress has extended the operating region to multigigabit/s speeds. Table 1 gives an overview of the state of the art in silicon ICs (adapted from [1]). Most circuit functions up to 10 Gbit/s can be implemented in silicon ICs. However, preamplifiers in this technology are still problematic because of the noise figure ; furthermore, more work is needed to obtain suitable high-frequency IC packages.

Two types of laser diodes are commonly used in transmission systems in the 1.3 μm or 1.55 μm wavelength region : Fabry-Perot (FP) lasers and Distributed Feedback (DFB) laser diodes. Degradations due to fibre dispersion effects are minimized by using DFB laser diodes with their very narrow optical emission spectrum, which makes them the preferred lasers in multigigabit/s systems [2]. Residual dispersion degradations are caused by chirp of the DFB laser, i.e. a small wavelength variation during modulation, and by spectral side modes which are insufficiently suppressed. Small-signal modulation bandwidths above 10 GHz have been reported, and large-signal modulation up to 16 Gbit/s [3].

There are also two options for selecting a photodiode operating in the 1.3 μm or 1.55 μm wavelength region [2] : the PIN photodiode, and the avalanche photodiode (APD) with its internal signal gain. Indium-gallium-arsenide (InGaAs) PIN photodiodes typically have a maximum bandwidth of 5 GHz, and in research 60 GHz has been achieved. Avalanche photodiodes (APDs) are available in InGaAs or germanium technology. The highest bandwidths and lowest excess noise are reached in InGaAs. Gain-bandwidth products of 75 GHz have been reported, and bandwidths up to 8 GHz.

Optical amplifiers (OAs) recently came into focus as having many advantages in high speed transmission systems. An OA converts incoming optical signals directly into amplified output optical signals, i.e. without converting them first to electrical signals like in conventional repeaters. It combines a high gain with low noise and a large bandwidth. Three application fields are of interest. As power booster directly after the transmitter laser, an OA increases the power going into the network. As in-line repeater it regenerates the signal intensity. Because of the OA's high signal transparency, the total network is easily upgradable to higher signal speeds or other signal formats. As a preamplifier at the receiver, the OA amplifies the signal directly before a PIN photodiode. Due to the OA's nearly unlimited bandwidth and its low noise, an OA-PIN photodiode combination outperforms in very high

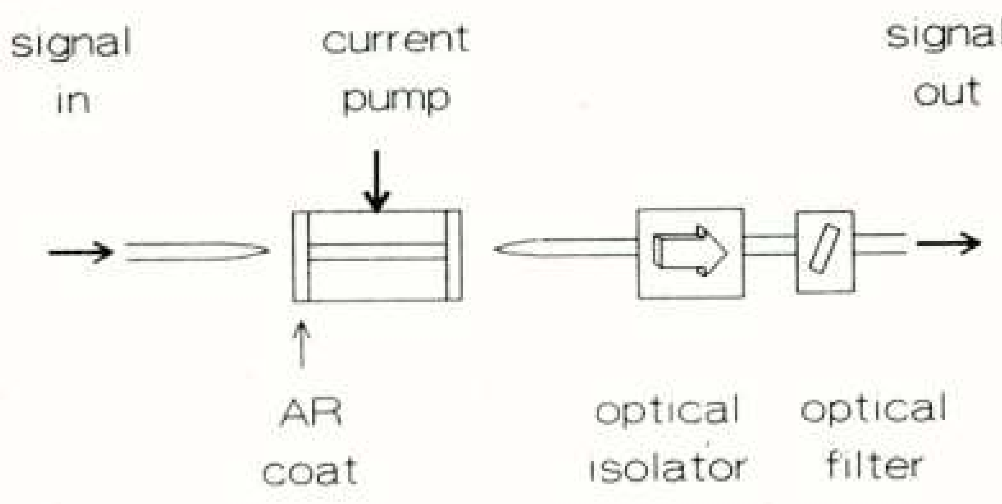


Fig. 3. Semiconductor laser amplifier

bitrate systems an APD, which has a limited gain-bandwidth product. OAs enable very long haul transmission, e.g. submarine links; they also enable a large variety in local network topologies, because they can elegantly compensate branching losses. Two main types of OAs can be distinguished : semiconductor laser amplifiers and fibre amplifiers.

Figure 3 shows the basic configuration of a semiconductor laser amplifier. It consists of a laser diode chip with highly efficient anti-reflection coatings on the facets in order to prevent laser action and to reduce the ripple in the gain amplitude versus frequency characteristics. Fibres are coupled to the device, and an optical isolator is used to avoid reflections. An optical bandpass filter can be included to reduce the spontaneous emission noise. A laser amplifier may operate in the 1.3 μm or the 1.55 μm region, and it has a large bandwidth of up to 70 nm. It can be integrated with optical waveguides in an opto-electronic integrated circuit. However, the polarization of the incoming signal should be matched to the laser amplifier. Fibre-to-device coupling losses are usually rather high (up to 5 dB), which reduces the net gain to 10 to 20 dB. Furthermore, the small time constant of the device may cause pattern effects if long strings of identical signal symbols are to be amplified.

Figure 4 shows the basic configuration of an erbium-doped fibre amplifier. The incoming signal is amplified by a fibre which is slightly doped with erbium. This material is brought into a phosphorizing state by pumping it with a high output power laser. Typical pumping wavelengths are 0.98 μm and 1.48 μm. The pump light is added to the signal with a wavelength division multiplexing (WDM) coupler. An optical isolator prevents reflection into the amplifier, and an optical bandpass filter reduces the spontaneous

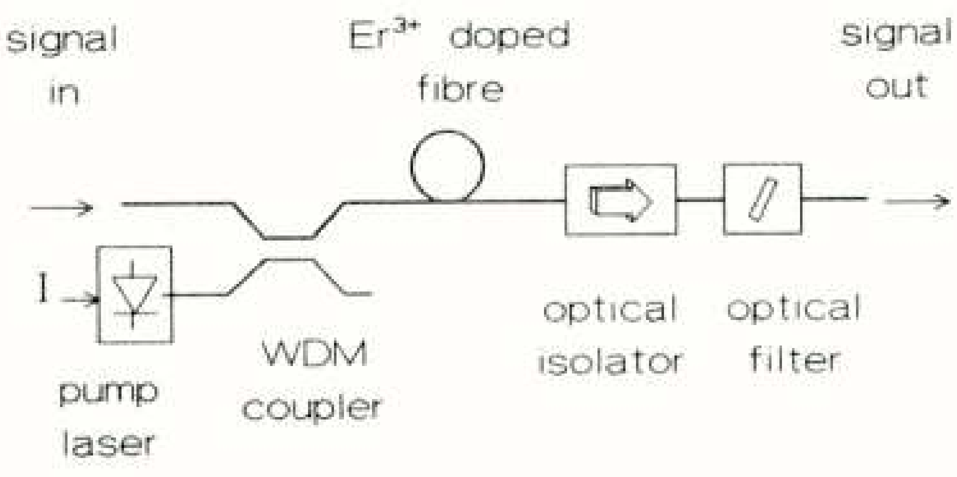


Fig. 4. Erbium-doped fibre amplifier

emission noise. An erbium-doped fibre amplifier operates on signal wavelengths around 1.54 μm. It may have a high net gain, up to 35 dB, because of its low coupling losses to the system fibre. It is insensitive to the signal polarization state, and it has no ripple in the passband characteristics. Crosstalk between wavelength multiplexed channels is very low. Because of its slow gain dynamics, signal pattern effects are negligible. The device is, however, not suitable for opto-electronic integration, and cannot operate at signal wavelengths in the 1.3 μm region. At high gain factors, its bandwidth around the 1.54 μm operating wavelength may be as small as 4 nm.

Optical power splitters located in the distant terminal of the passive tree-and-branch network distribute the signal from the feeder line over the individual subscriber drop lines. They can be composed with elementary 2x2 couplers as illustrated by figure 5. A $1{:}2^M$ splitter has a splitting loss α_{spl} given by

$$\alpha_{spl} = M\ (\ 3 + \eta + \alpha_s\ ) - \alpha_s \qquad [dB]$$

where η denotes the excess loss of a 2x2 coupler, and α_s the loss of a fibre splice in dB.

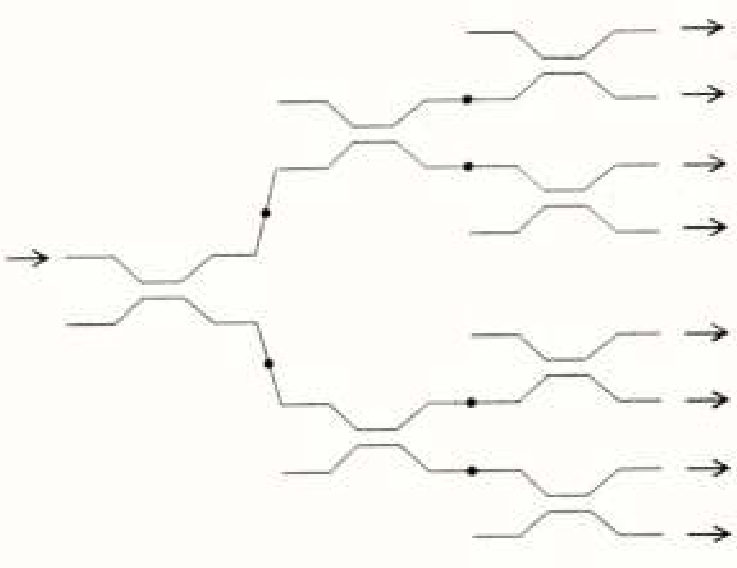


Fig. 5. Optical splitter, composed of 2x2 couplers

4. <u>System limits</u>

The sensitivity of an optical receiver degrades with increasing signal bitrate. Figure 6 shows the average received power at a signal wavelength of 1.5 μm, needed for a bit error rate of 10^{-9} and calculated for typical values of the receiver parameters. Clearly an APD photodiode in combination with a GaAs FET preamplifier yields a better receiver sensitivity than a PIN-FET receiver. However, this advantage diminishes at higher bitrates due to the limited gain-bandwidth product of the APD. At very high bitrates the best receiver sensitivity is reached by an optical preamplifier in front of a PIN-FET receiver, although practical results up to now are still quite far from the theoretical limit.

In the network configuration shown in figure 1, the number of subscribers which can be fed from a single

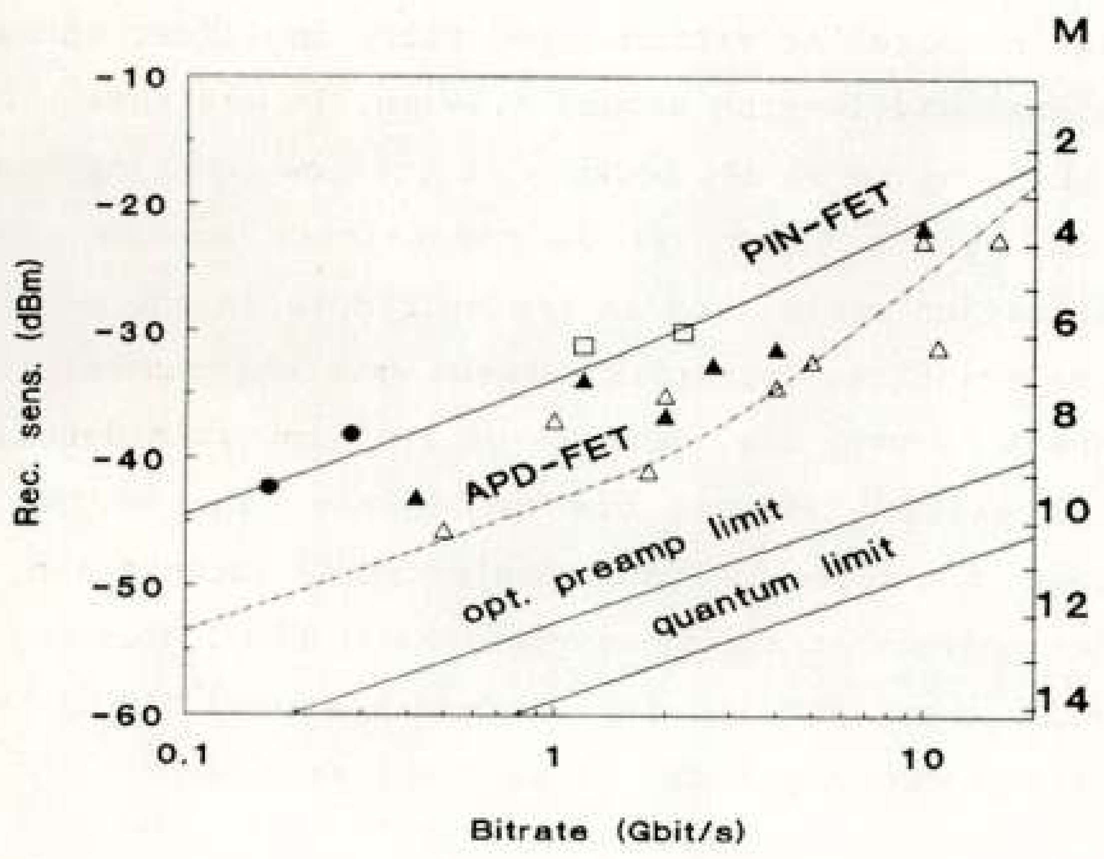


Fig. 6. Sensitivity of direct detection receiver at λ=1.5 µm ; splitting factor 2^M for 6 km feeder fibre, 1 mW transmitter output power, and negligible dispersion
(theory : -·- InGaAs APD, G*B=20 GHz.
experimental results : □ Ge APD, λ=1.3 µm;
▲ InGaAs APD; ● InGaAs PIN, λ=1.3 µm;
△ optical preamplifier)

feeder fibre is determined by the fibre attenuation and by the power budget between transmitter and receiver ; it is assumed that the fibre link is short enough to neglect dispersion effects. Figure 6 shows the number of subscribers 2^M where M is given as a function of the bitrate. A feeder fibre length of 6 km and an output power level for the distributive services of 1 mW at the input of the OLI has been assumed. The elementary 2x2 couplers in the optical power splitter are assumed to have an excess loss of 0.5 dB, and the connecting splices a loss of 0.2 dB. At 10 Gbit/s 16 subscribers can be fed if an APD-FET receiver is used ; with an optical preamplifier in front of a PIN-FET receiver more than 64 subscribers can be fed.

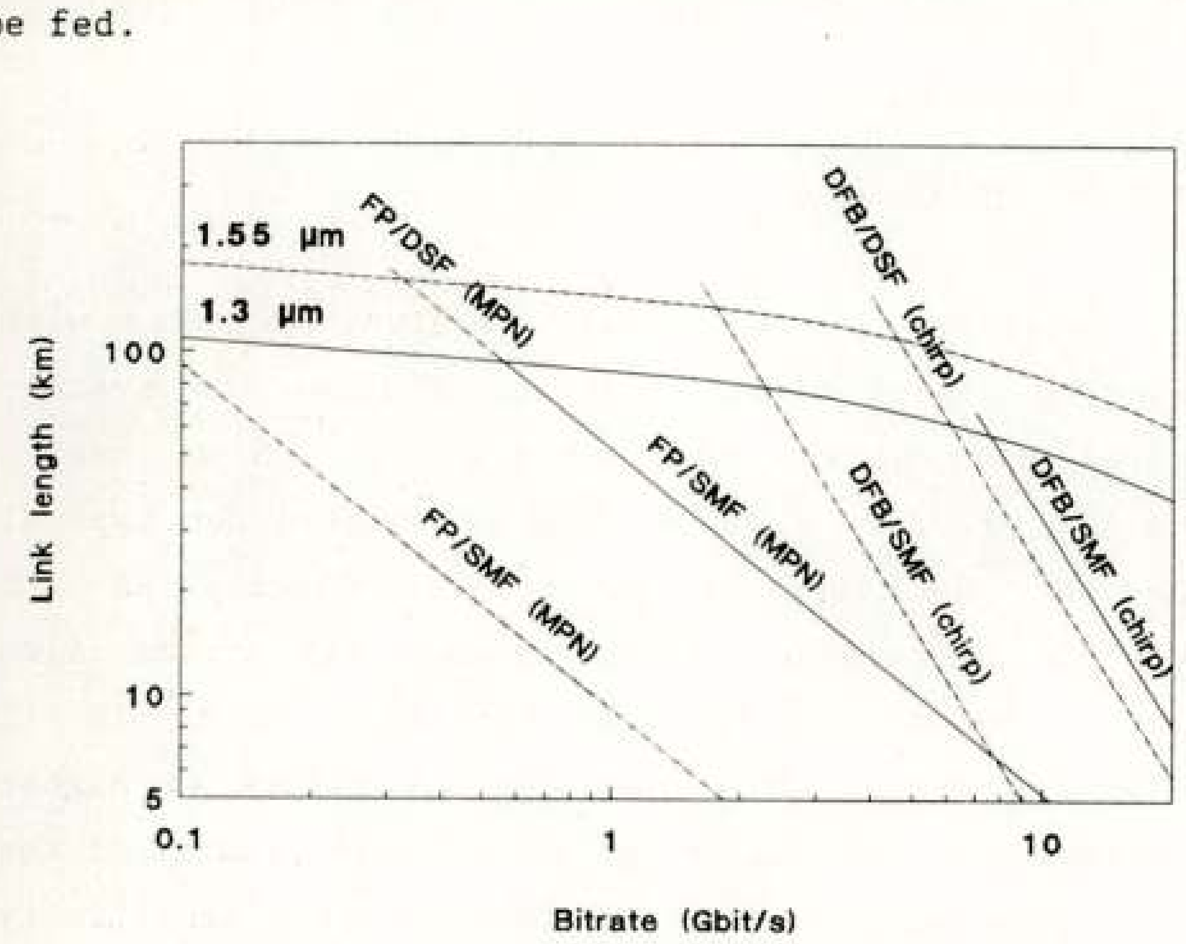


Fig. 7. Limits of direct detection transmission systems on single mode fibre
(FP laser : spectral width $\Delta\lambda$=4 nm FWHM, MPN k=0.15; DFB laser : linewidth enhancement factor α=4; InGaAs APD ; —— λ=1.3 µm ; -·- λ=1.55 µm)

The feeder fibre length is not only limited by the available system power budget, but also by dispersion effects. The maximum fibre link length versus bitrate is shown by figure 7 for various types of fibres and laser diodes. The nearly horizontal curves represent the attenuation-limited link lengths at 1.3 µm and 1.55 µm. The other curves show the dispersion-limited link lengths assuming a dispersion penalty of 1 dB. The limiting dispersion mechanisms are mode partition noise (MPN) occurring in FP laser diodes, and chirp occurring in DFB laser diodes (see [2]). Clearly, to achieve 10 Gbit/s transmission at 1.55 µm wavelength over feeder links exceeding 5 km, a DFB laser has to be used in combination with a dispersion shifted single mode fibre (DSF, having zero dispersion at about 1.55 µm). However, standard single mode fibre (SMF, having zero dispersion at about 1.3 µm) probably remains cheaper and is being installed in large volume presently. Very high bitrates like 10 Gbit/s can only be transported over reasonable lengths of SMF if a low-chirp transmitter modulation method is adopted.

A low chirp transmitter may consist of a DFB laser diode operating at a constant optical output power, followed by an optical modulator. An electro-absorption modulator can be monolithically integrated with the DFB laser ; recently a device operating at 10 Gbit/s has been reported, operating in a system with 65 km dispersion-shifted fibre [4]. However, such a device still exhibits a small chirp. Another modulator type is the Mach-Zehnder interferometer, shown in figure 8, where the intensity modulation is brought about by modulating the phase of the light signal in each of two parallel waveguides and thus controlling the amount of constructive interference at the point where the waveguides join again. The chirp with a Mach-Zehnder modulator is negligible provided that the phase modulators are driven with opposite signal phases. However, the modulation signals needed are rather high (typically 5 V) and therefore hard to generate at very high bitrates. Insertion losses are also in the order of 2 dB or higher. Bandwidths exceeding 5 GHz have been reported [5].

An alternative approach to realize a low chirp transmitter is to modulate the frequency (or phase) of the DFB laser, and convert this into an amplitude modulation signal with an external interferometer. The basic

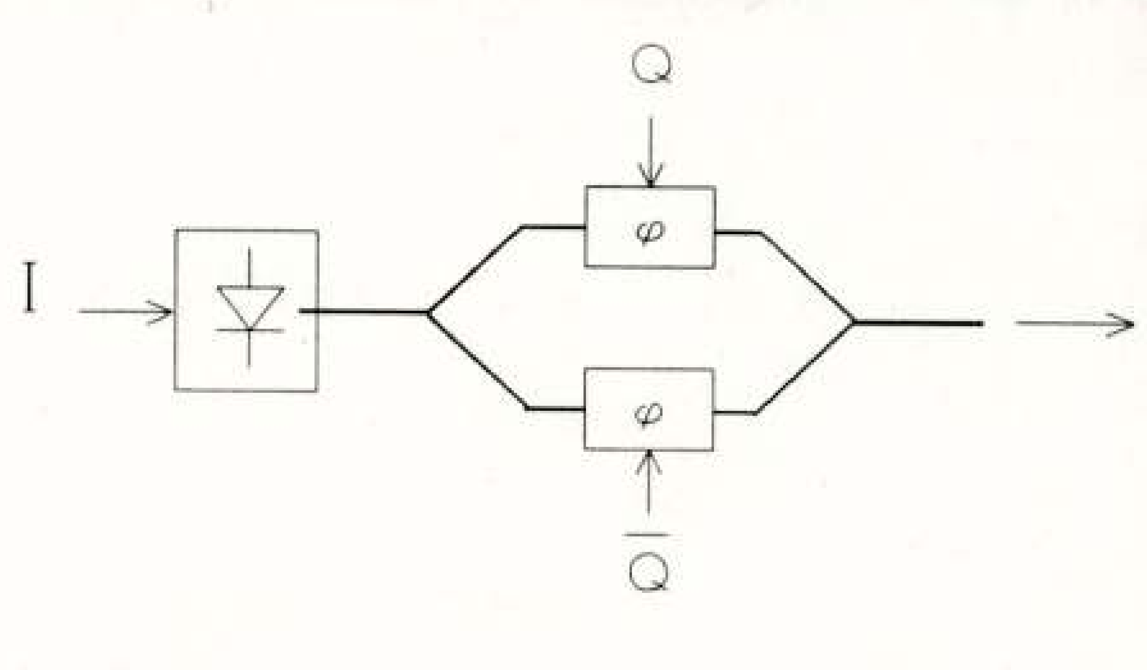


Fig. 8. Mach-Zehnder interferometer

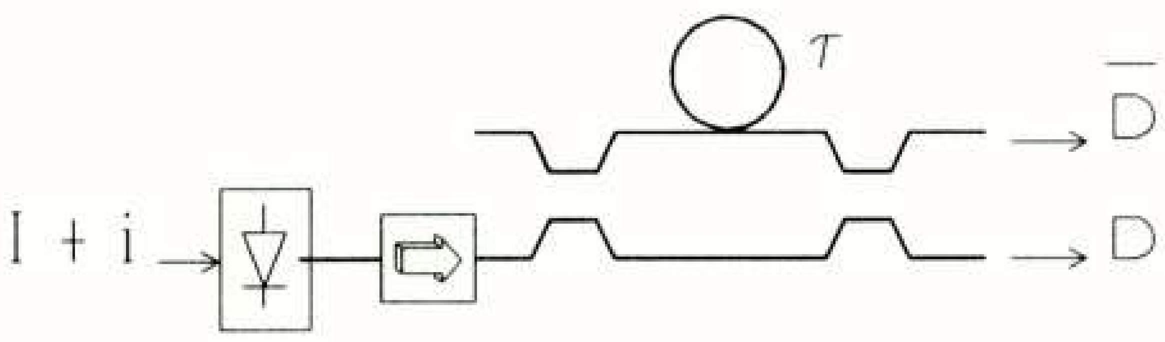


Fig. 9. Phase (or frequency) to intensity modulation conversion

configuration is shown in figure 9. Modulating the DFB's frequency is achieved simply by adding a small signal current to the laser bias current. As only a small current swing is needed, modulation at high bitrates is considerably facilitated. However, the laser should have a narrow spectral linewidth, a flat FM response characteristic, and its emission frequency should be locked to the characteristic of the interferometer. Recently a 20 Gbit/s 1.55 μm transmitter with phase modulation of the laser has been reported, operating in a system which could accommodate dispersions up to 330 ps/nm (i.e. about 18 km SMF) [6].

5. RACE-project 1051 : Multigigabit/s transmission in the IBCN subscriber loop

The EC-funded RACE program aims at the installation of an Integrated Broadband Communication Network (IBCN) all over Europe, starting in 1995. Both interactive and distributive services should be integrated in a single network, Several optical fibre transmission techniques are proposed for the distribution of broadband services, such as analog modulation using subcarrier multiplexing , coherent multichannel transmission, and multigigabit/s direct detection techniques.

The RACE project 1051 is concentrating on 5 to 10 Gbit/s transmission with direct detection, to be applied in TV or HDTV feeder and distribution networks. The partners involved are SEL (prime contractor), ANT, APT, FI-FTZ, GPT, IST, NKT, Telefonica and University of Athens. In 1992, at the end of this project, a laboratory set-up must be completed which demonstrates an optical distributive network carrying 10 Gbit/s traffic. The electronic functions should all be built into ICs which are preferably made in silicon technology. The first experimental results are quite promising.

The network brings a single fibre to every subscriber transporting both interactive and distributive traffic. The customer premises network (CPN) may distribute the broadband signals in-house via electrical means (fibre-to-the-home, FTTH). In this case the network termination contains the high-speed optical receiver and a channel selector, which extracts the selected services out of the high bitrate stream and feeds them to the videosets via a coaxial CPN. The CPN may also distribute the signals via optical fibres (fibre-to-the-terminal, FTTT). Then the network termination only contains an optical splitter, which distributes the high bitrate stream optically to the videosets. Every set should have a high speed optical receiver and a channel selector. Both FTTH and FTTT concepts are being studied.

6. Conclusions

There is a clear and increasing need for the transportation of very high bitrates in local loop networks. The distribution of video signals, in particular of the near-future HDTV signals, is the main driving force requiring multigigabit/s data speeds fed via optical fibre to the residential subscribers.

Fully passive optical local networks offer the most future-proof infrastructure for broadband communications. They allow an evolution in data transport techniques while leaving the network itself unchanged. This evolution may start on short notice with subcarrier multiplexing techniques. Multigigabit/s direct detection techniques offer a higher potential in terms of number of subscribers and signal quality. The ultimate potential is given by coherent multichannel techniques.

Within the RACE program, the aforementioned three techniques are all being studied. The RACE 1051 project is concentrating on 5 to 10 Gbit/s direct detection systems. It is expected that nearly all high speed functions can be realized in silicon ICs, which holds good promise for economic local networks. In research various laboratory set-ups have already demonstrated that direct detection local loop systems with speeds exceeding 10 Gbit/s are technically feasible using potentially low-cost (opto-)electronic components.

References

[1] H.-M. Rein : "Silicon bipolar integrated circuits for multigigabit per second lightwave communications", Proc. of Opt. Fiber Comm. Conf., San Francisco, 22-26 Jan. 1990, paper WF-2

[2] A.M.J. Koonen : "Fibre-optic components for multigigabit/s transmission systems", Tijdschrift van het NERG, Vol. 53, No. 3, 1988, pp. 95-99

[3] A.H. Gnauck et al. : "16 Gbit/s transmission experiments using a directly modulated 1.3 μm DFB laser", IEEE Phot. Techn. Lett., Vol. 1, No. 10, Oct. 1989, pp. 337-339

[4] T. Okiyama et al. : "A 10-Gb/s, 65-km optical fiber transmission experiment using a monolithic electro-absorption modulator/DFB laser light source", Proc. of 15th Eur. Conf. on Opt. Comm., Gothenburg, 10-14 Sept. 1989, paper MoA1-3

[5] T. Namiki et al. : "Low-drive-voltage Ti:$LiNbO_3$ Mach-Zehnder modulator using a coupled line", Proc.

of Opt. Fiber Comm. Conf., San Francisco, 22-26 Jan. 1990, paper TUH-4

[6] M. Shirasaki et al. : "20 Gbit/s no-chirp intensity modulation by DPSH-IM method and its fiber transmission through 330 ps/nm dispersion", Proc. of 15th Eur. Conf. on Opt. Comm., Gothenburg, 10-14 Sept. 1989, paper PDA-1

Voordracht gehouden tijdens de 376e werkvergadering.

REAL-TIME VIDEO, VAN IDEE TOT IC

Dr. A.H.M. van Roermund
Philips Natuurkundig Laboratorium

Real-time video, from idea to IC. IC-design methods are discussed in the context of real-time video; a general-purpose approach based on a programmable processor is further detailed.

Implementatie van real-time video-algoritmen in IC's is een moeilijke zaak; de daaraan voorafgaande simulatie en hardware prototyping al evenzeer. Van alle kanten wordt naarstig gezocht naar oplossingen. Verschillende wegen beginnen zich af te tekenen. De vraag waar de ontwerper zich nu voor gesteld ziet is: wat is de beste weg?

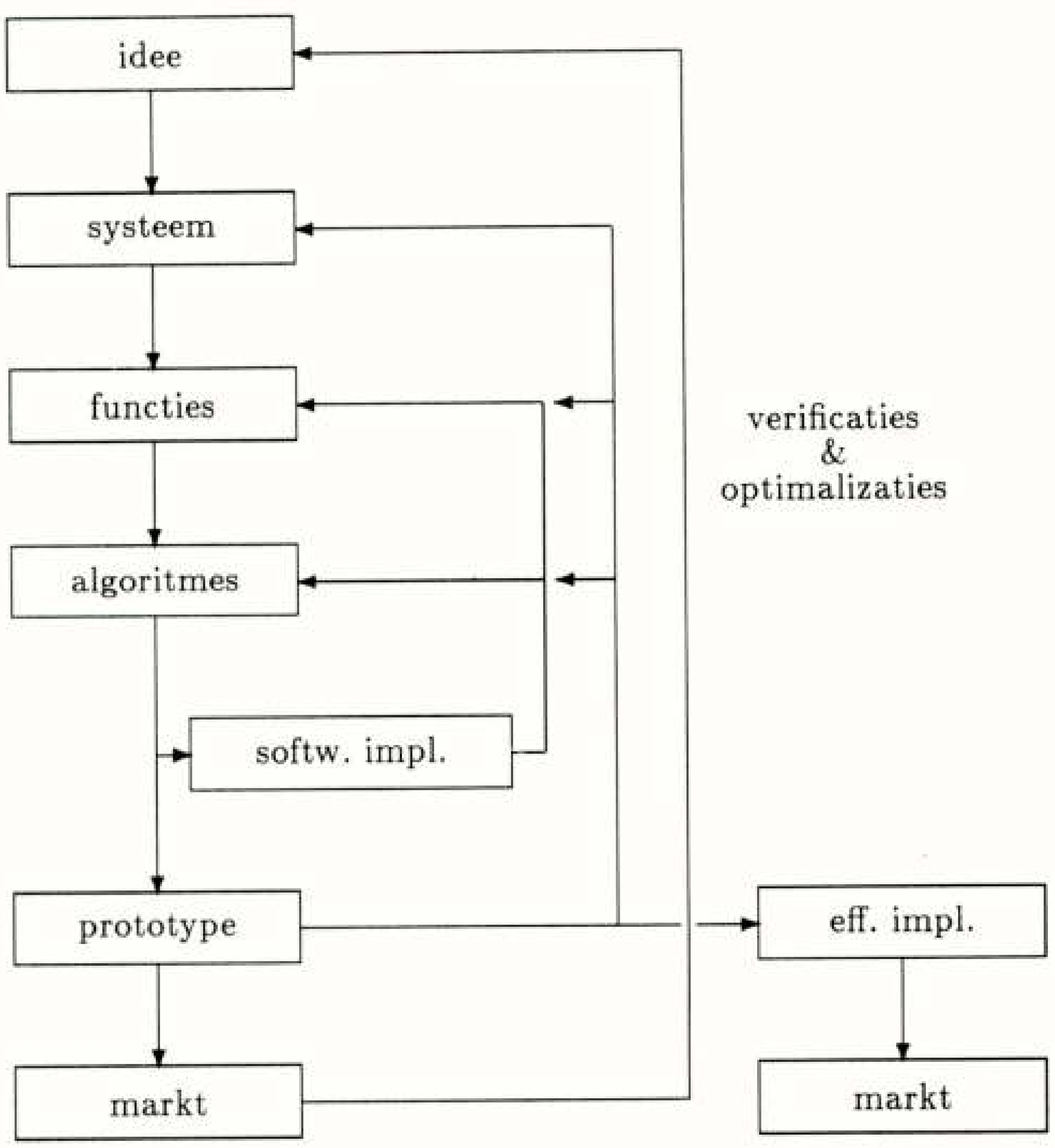
In de kern van de zaak ligt het probleem bij de hoge datasnelheden bij video en de daarmee gepaard gaande eisen t.a.v. processing-power. Bij data- of audioprocessing ligt dat heel anders.

In dit artikel zal eerst iets uitvoeriger worden ingegaan op de bovengeschetste probleemstelling; daarna zullen de verschillende opties voor IC-implementatie ten behoeve van een vroegtijdige verificatie en optimalisatie van het algoritme en van een efficiënte implementatie van het algoritme op silicium besproken worden. Duidelijk zal worden gemaakt, dat er niet zoiets als een universele 'beste' oplossing is, maar dat de ontwerper op grond van zijn specifieke eisen een geschikte keuze moet maken. Een nieuwe implementatiewijze, gebaseerd op een general-purpose programmeerbare processor-chip, zal verder worden uitgewerkt.

1 Van idee tot IC

Het trajekt 'van idee tot IC' is gevisualiseerd in Fig 1. Uitgaande van een aantal ideeën omtrent toekomstige produkten zal een globaal systeemconcept worden gespecificeerd. Een verdere uitwerking vindt plaats door het concept te vertalen in een combinatie van een aantal deelfuncties, en vervolgens bij elke deelfunctie algoritmes te zoeken die deze functies mogelijk maken.

Deze voorstelling suggereert een eenduidige ontwerpvolgorde, van boven naar beneden (*top down*-ontwerp). In de praktijk is dat niet zo. Enerzijds komt dat omdat de specificaties niet volledig zijn: de systemen waar het om gaat zijn niet volledig bekend en de onderdelen die wel bekend zijn zijn in de meeste gevallen niet formeel beschreven wegens gebrek aan beschrijvingswijzen of omdat er nog niet voldoende aandacht aan geschonken is. Daarnaast spelen bij het opstellen van de specificaties de uiteindelijke hardware-aspecten een duidelijke rol en die weet je vooraf niet volledig; je werkt in de praktijk met afschattingen.



Figuur 1: Ontwerptrajekt, van idee tot IC

Naast het verificatieprobleem hebben we bovendien te maken met een optimalisatieprobleem: voor elke afbeeldingsslag in een ontwerptrajekt bestaan in principe veel mogelijkheden ('ontwerpboom'), waaruit de ontwerper een optimale keuze moet proberen te maken. Bijvoorbeeld: bij het vastleggen van een specificatie voor datacompressie staan er een aantal compressiemethodieken ter beschikking; een gevraagde filtercurve

kan met verschillende filters gerealiseerd worden; operaties kunnen via berekening worden geïmplementeerd of via opzoektabellen; er is keuze in volgorde van bewerkingen, in parallel werken of serieel of m.b.v. een *pipe line*-structuur, en in de wijze waarop geheugens worden gedistribueerd; etc. De consequentie is dat de resultaten van verschillende keuzes moeten worden geverifiëerd en vergeleken ten behoeve van optimalisatie.

Verificatie en optimalisatie gebeuren in eerste instantie met computersimulaties op het laagste, algoritmische niveau en vandaaruit steeds een stap hoger ('bottom up'). Wanneer men al die slagen heeft voltooid en een (voorlopige) keuze heeft bepaald, volgt een gelijksoortige aktie, maar nu op hardware-niveau; de computersimulaties zijn namelijk beperkt naar modelvorming, snelheid en rekenkracht. Met deze hardware, ook wel prototype genoemd, kan bovendien geverifiëerd worden of de oorspronkelijke doelstellingen en specificaties inderdaad in overeenstemming zijn met wat de márkt vraagt (uitzetten van een aantal systemen; demonstraties). Zonodig worden reeds eerder genomen beslissingen bijgesteld.

Op dit moment is het ontwerp in principe af en volgt de produktontwikkeling, waarbij de efficiëntie van de implementatie belangrijk is. Efficiëntie kan daarbij vertaald worden naar bijvoorbeeld snelheid als het om *pilot*-produktie gaat, minimale hardware als het om grote aantallen gaat, of flexibiliteit als het om een produkt gaat dat een zekere flexibiliteit vraagt.

Bezien we nu opnieuw het totaal-ontwerptrajekt van Fig. 1, dan zien we dus op twee punten IC-implementatie om de hoek kijken: bij de prototyping, dus omwille van verificatie en optimalisatie in de ontwikkelingsfase van een produkt (algoritmeontwikkelling), en bij de uiteindelijke produktimplementatie.

2 Knelpunten

In het voorgaande hoofdstuk is het ontwerptrajekt van idee tot IC in zijn algemeenheid beschreven. De titel van het verhaal luidt echter: '*Real-time video*, van idee tot IC'. Wat maakt real-time video nu zo specifiek en wat betekent dat voor dit ontwerptrajekt?

Specifieke aspecten van real-time video

Real-time video betekent: een *continue* stroom van tweedimensionale (kleuren)beelden. Stel het aantal beeldpunten ('pixels') op 400.000 en het aantal benodigde bits per pixel op 12 (± huidige TV-norm). Dat betekent per beeld een informatieinhoud van zo'n 5 Mbit. Bij een beeldherhalingsfrequentie van 25 Hz (TV) komt dit neer op een datastroom van zo'n 125 Mbit/s. Nemen we nog een aantal andere factoren mee zoals horizontale en verticale overscan (lijn- en beeldterugslag), dan komen we al op meer dan 160 Mbit/s. De tendens is bovendien naar hogere resolutie; bij 'HDTV-kwaliteit' resulteert dat in een datastroom van al gauw meer dan 1 Gbit/s.

Voor opslag en transmissie wordt daarom intensief gezocht naar compressietechnieken en bij processing ziet men zich gesteld voor een enorme eis t.a.v. rekenkracht en datastroom.

Een ander specifiek aspekt van videoprocessing is dat de meerdimensionale informatie in de meeste gevallen is versleuteld in één videosignaal (scan-domain). Dit betekent dat we naast processing op luminantie- en chrominantiesignalen ook processing moeten doen op synchronisatie- en controlsignalen en op combinaties ervan. Elk van die type signalen wordt in de praktijk op verschillende manieren behandeld; zo is de bemonsteringfrequentie voor deze signalen bijvoorbeeld verschillend.

Tenslotte kunnen we nog opmerken dat videobewerkingen meestal plaatsvinden in het drie-dimensionale beeldvlak/tijddomein en daarmee, op grond van de bovengenoemde versleuteling van de beeldinformatie in een ééndimensionaal signaal, veel geheugenfuncties vragen en dus pixel-, lijn-, raster- en beeldvertragingen.

Consequenties voor het ontwerptrajekt

De belangrijkste consequenties van het bovenstaande voor het ontwerptrajekt zijn dat de computersimulaties beperkt zijn, dat prototyping veel mankracht en tijd kost en vaak leidt tot onhandelbaar grote hardwaresystemen en dat de specificaties voor de IC-implementaties aan de rand zitten van wat haalbaar is, gezien de technologische grenzen aan snelheden, dissipatie en oppervlak. Dat maakt een goede keuze op elk niveau in het ontwerptrajekt van essentiëel belang.

De computersimulaties vragen een computer met grote rekenkracht en snelle I/O hardware (disks, RAMs, scan conversies) waarmee op volle snelheid sequenties van videobeelden kunnen worden ingenomen en, na bewerking, weergegeven. Dat is een dure aangelegenheid. Noodgedwongen wordt het aantal sequenties minimaal gehouden en ook de lengte per sequence wordt kort gehouden (±4 sec.). Gezien het perceptieve en daarmee subjectieve aspect van videoprocessing is het zeer de vraag of dergelijke sequenties wel voldoende representatief en maatgevend zijn. Daarenboven vraagt optimalisatie van een systeem, in tegenstelling tot verifi-

catie, méérdere simulaties, telkens met een andere set van waarden voor de systeemparameters. Aangezien elke simulatie veel tijd kost (vaak in de orde van uren), worden optimalisaties slechts in beperkte mate op dit niveau uitgevoerd; in de praktijk worden ze uitgesteld tot in de hardware-fase.

Prototyping en uiteindelijke implementatie zijn ook complex en tijdintensief, met als gevolg dat het pad 'van idee tot IC' te lang, te kostbaar en te risico-vol is. Vanuit dat oogpunt bezien is er een behoefte aan flexibele hardware waarmee zowel de prototyping als de produktimplementatie kan worden versneld.

3 Vele wegen naar Rome?

De oplossing moet dus gezocht worden in het vereenvoudigen en versnellen van het IC-ontwerptrajekt. Meerdere ontwikkelingen zijn momenteel zichtbaar in de IC-ontwerp-wereld, die in deze richting gaan. Zijn er dus vele wegen die naar Rome leiden? In zijn algemeenheid kunnen we zeggen dat versnelling in het IC-ontwerptrajekt wordt bereikt door het aantal vrijheden in te dammen, anders gezegd: door te veralgemeniseren. Dat betekent echter, dat voor het merendeel van de gevallen niet meer de optimale prestatie kan worden bereikt die met *full custom*-ontwerp mogelijk is. En dat is bij video nu net problematisch omdat de marges al zo klein zijn; we zitten vaak al dicht bij de rand van de technologische mogelijkheden.

3.1 Implementatiemethodieken

IC-implementatiemethodieken kunnen op verschillende manieren geclassificeerd worden. De classificatie die hier wordt gekozen is gebaseerd op de manier waarop gebruik wordt gemaakt van vooraf vastgelegde keuzes.

Full-custom design

Full-custom design kan gezien worden als het ene uiterste: niets veralgemeniseren; alles specifiek voor de gegeven situatie; optimaal naar uiteindelijke prestaties, maar tijdrovend en dus resulterend in hoge ontwerpkosten. Voor min of meer standaard functies, dus grote oplagen, kan dit een goede methode zijn.

Vaste basisfuncties

De cell-based en standard-cell methoden gaan uit van vaste basisfuncties, opgeslagen in bibliotheken, die kunnen worden opgeroepen en gecombineerd teneinde tot de gewenste totaalfunctie te komen. Het IC-fabricagetrajekt wordt nog steeds volledig doorlopen; alle maskers worden 'customer-specific' gemaakt.

Vaste basisstructuren

Een andere benadering is dat niet de functie, maar de geometrische structuren gedeeltelijk worden vastgelegd; dit vinden we bijvoorbeeld bij de gate-array en cell-array methoden. Een aantal processtappen kan bij voorbaat worden uitgevoerd. Slechts de laatste maskerstap(pen) maken het ontwerp specifiek. Deze methode wordt vaak gecombineerd met de eerdergenoemde methode: functies zijn vantevoren afgebeeld naar vaste basisstructuren en als basisfunctie opgeslagen in bibliotheken. Een goed uitgewerkte bibliotheek samen met bijvoorbeeld een *high-density gate array* kan tot krachtige implementaties leiden.

Programmeerbare structuren

In principe wordt bij de programmeerbare structuren de *hele* structuur vastgelegd, en wordt de gebruiker in de gelegenheid gesteld om achteraf, via programmeren de IC-functie gebruikers-specifiek te maken. Dit programmeren kan op verschillende niveau's gebeuren: op gate-, cell, of zelfs op processor-niveau. Maskerprogrammeerbare IC's kunnen worden gezien als een tussenvorm tussen deze en voorgaande methode.

Geparametrizeerde functies

Uitgaan van vaste basisfuncties gaat goed zolang de functies nog van een redelijk laag niveau zijn; bij complexere functies worden de functies veel te specifiek en daarmee het toepassingsgebied ervan te klein. Daarom wordt in dat geval gebruik gemaakt van (procedurele) geparametrizeerde functies, die als zogenaamde *macro's* worden opgeslagen. Vandaar ook de naam: macro-blokmethode.

Vaste basisarchitectuur

Bij een IC-ontwerp is het ook mogelijk om uit te gaan van een bepaald architectuurmodel. Daarbij kan bijvoorbeeld gedacht worden aan het vastleggen van het type datapad, de communicatiestructuur (buslijnen met protocol, vaste communicatielijnen, ...), etc. Op deze manier worden dan de vrijheden van het ontwerptrajekt ingeperkt. Het applikatie-specifiek maken van de architectuur kan daarbij in meer of mindere mate geautomatiseerd worden, al dan niet op basis van heuristieken of we kennis -banken; we spreken dan meestal van silicon compilers. Het IC-trajekt kan zo sterk versneld worden; het IC-fabricagetrajekt (processing) moet daarentegen nog steeds volledig worden doorlopen.

3.2 Geschiktheid voor video

Zoals reeds opgemerkt, is het probleem bij IC-implementaties t.b.v. video processing dat de eisen t.a.v. snelheid, dissipatie en rekenkracht zó stringent zijn dat niet teveel water in de wijn mag worden gedaan bij het uitgaan van vaste functies, structuren, architecturen of ontwerpkeuzes. Per specifiek geval zal nu gekeken moeten worden wat de beste keuze is; elke toepassing zal namelijk zijn eigen gewichten toekennen aan de specifieke voor- en nadelen van elk van de eerder genoemde methoden. Bovendien zijn lang niet alle methoden ver genoeg uitgewerkt voor video: compilers voor videotoepassingen bevinden zich bijvoorbeeld nog in de research-fase.

Een aantal belangrijke beoordelingspunten, op grond waarvan een keuze moet worden gemaakt tussen de ontwerpmethodieken, zijn:

- elektrische prestaties
- IC-oppervlak en IC-kosten
- ontwerpflexibiliteit
- ontwerpkosten en -tijd
- risico's in het ontwerptrajekt

Het relatieve belang van deze criteria wordt bepaald door de aan het produkt gestelde eisen, vanuit de markt gezien, zoals:

- aantallen
- prijsdruk
- levensduur
- huidige ontwikkelingsfase
- functionaliteit
- flexibiliteit

Ook vanuit de ontwerp- en produktiecapaciteit gezien worden eisen gesteld. Zo zijn bijvoorbeeld het aantal beschikbare ontwerpers en de aanwezige expertise en software beperkende factoren. Een kleine firma kan zich bijvoorbeeld niet veroorloven om expertise en software op het gebied van alle eerdergenoemde methoden in huis te halen.

Ter verduidelijking worden een aantal voorbeelden gegeven. Hierbij zullen laatstgenoemde beperkingen t.a.v. ontwerp- en produktiecapaciteit buiten beschouwing worden gelaten.

Voor *standaardfuncties in een TV-ontvanger*, zoals bijvoorbeeld een kleurendecoder, zijn grote aantallen, lage prijs en gestabiliseerde functie (en dus lange levensduur) kenmerkend. Dientengevolge zijn de kosten en de duur van het ontwerptrajekt niet zo belangrijk en evenmin de flexibiliteit van het produkt. De kosten van het uiteindelijk IC daarentegen (en dus IC-oppervlak) zijn zeer belangrijk. Voor dergelijke gevallen zal men dus al gauw kiezen voor full-custom design.

Features in een TV zijn functioneel gezien over het algemeen nog aan veranderingen onderhevig, en meer *high end*-gericht, voor kleinere aantallen. Met name voor nieuwe features is een IC-ontwikkeling dus al gauw relatief duur. Een snel en goedkoop ontwerptrajekt en flexibiliteit is gewenst. Een gate array kan in veel gevallen een goede oplossing zijn, later eventueel gevolgd door een full-custom ontwerp zodra de functie zich stabiliseert en de aantallen beginnen op te lopen. Een programmeerbare oplossing kan de gewenste flexibiliteit leveren.

De *multi media*-markt is een opkomende markt, waarbij diverse soorten media (audio, video, data, tekst, ...) samenkomen in één systeem. Hiertoe behoren CDI (Compact Disk Interactive) en Desk-top Video (werkstations of PC's met uitbreiding naar full-motion video). Dit zijn produkten die nog aan het begin van hun ontwikkeling staan, met kleine aantallen, met nog niet gestabiliseerde specificaties, en met een hoge eis t.a.v. interaktiviteit van de gebruiker. Gezien bovendien de korte economische levensduur van IC's in bijvoorbeeld werkstations en PC's ($1\frac{1}{2}$ á 2 jaar) zijn snelheid van ontwerp, flexibiliteit en programmeerbaarheid hier belangrijke factoren.

Professionele produkten worden gekenmerkt door kleine aantallen, hoge prijzen. De kosten van een IC zijn hier niet zo belangrijk als bij consumentenprodukten; snelheid van ontwerp, flexibiliteit en mogelijk programmeerbaarheid van het produkt daarentegen wegen hier zwaarder.

Bij *prototyping* hebben we per definitie te maken met weinig prijsdruk, korte levensduur en onzekerheid omtrent juistheid (of zinvolheid) van de geïmplementeerde functie. Belangrijk zijn hier: flexibiliteit en ontwerpsnelheid.

4 De VSP-aanpak

Bij Philips wordt gewerkt aan verschillende van de eerder genoemde IC-implementatiemethoden. Eén ontwikkeling, waar hier verder op ingegaan zal worden, is die van een programmeerbare (digitale) videosignaalprocessor (VSP). Hiermee kunnen op een snelle en flexibele wijze real-time videoapplikaties worden gerealiseerd. Processoren voor dataprocessing, telecommunicatie en signaalprocessing zijn al lang bekend en volop in gebruik; voor video-applikaties is deze aanpak nieuw.

Als voordelen kunnen worden genoemd:

- snel algoritmeontwerp;
- grote flexibiliteit;
- goede optimalisatie door snelle iteraties;
- geen IC-ontwikkelingskosten;
- werking van IC's (VSP's) bij voorbaat gegarandeerd, dus risico's van IC-trajekt vermeden;
- extra functionaliteit in de vorm van programmeerbaarheid.

Gezien de voorgaande discussie zal het duidelijk zijn, dat deze aanpak met name geschikt is binnen een ontwikkelomgeving, voor prototyping en voor produktimplementatie in de professionele en multi-media-sfeer.

De VSP is gerealiseerd en volledig functioneel. Tegelijkertijd is gewerkt aan software voor de ondersteuning bij het programmeren (*support tools*). Het resultaat is een gebruikersvriendelijk pakket waarmee binnen een tijdsbestek van een paar uur tot een paar dagen volledige applikaties kunnen worden geprogrammeerd ('afgebeeld'). Verschillende afbeeldingen zijn reeds gemaakt. Fig. 2 toont een typische totaalopstelling: een board met 3 VSP's; een monitor waarop het uitgangssignaal na bewerking door de VSP's wordt weergegeven (in dit geval een door één VSP gegenereerd testbeeld); een Apollo werkstation waarmee met behulp van de programmeersoftware het programma voor de VSP('s) wordt ontwikkeld.



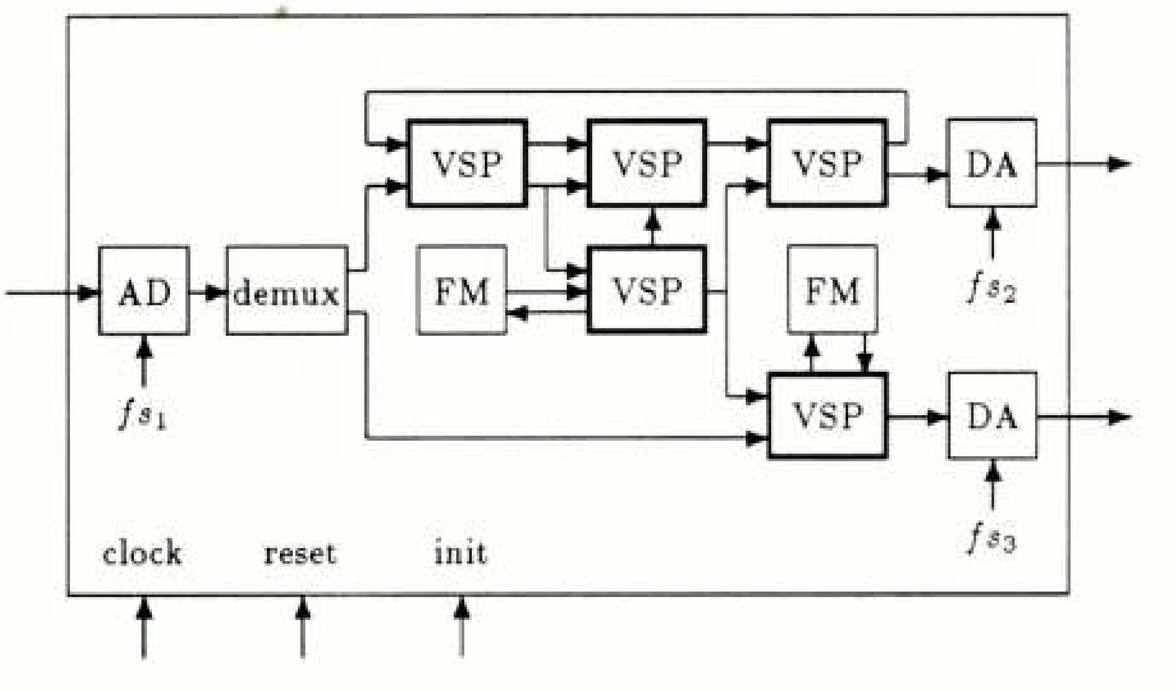
Figuur 2: Een typische VSP-omgeving

De architectuur van de VSP, de opbouw van de tools en de wijze van afbeelden worden elders in de literatuur al uitvoerig beschreven (bijvoorbeeld in [1]); hier zullen we ons beperken tot een globale, en daardoor soms wat compacte, beschrijving.

VSP-kenmerken

Bij de definitie van de VSP is uitgegaan van een aantal keuzes, die in grote mate de architectuur hebben bepaald en die daarmee ook kenmerkend zijn voor de VSP:

- general-purpose binnen het domein van real-time video;
- multi-processor systeemopbouw;
- goede balans tussen communicatie- en processingeigenschappen;
- interne multiplexing voor optimaal gebruik van processing- en communicatie-hardware;
- gelijktijdige processing op alle type signalen;
- gelijktijdig operaties op verschillende snelheden;
- volledig synchrone werking;
- eenvoudige programmering.

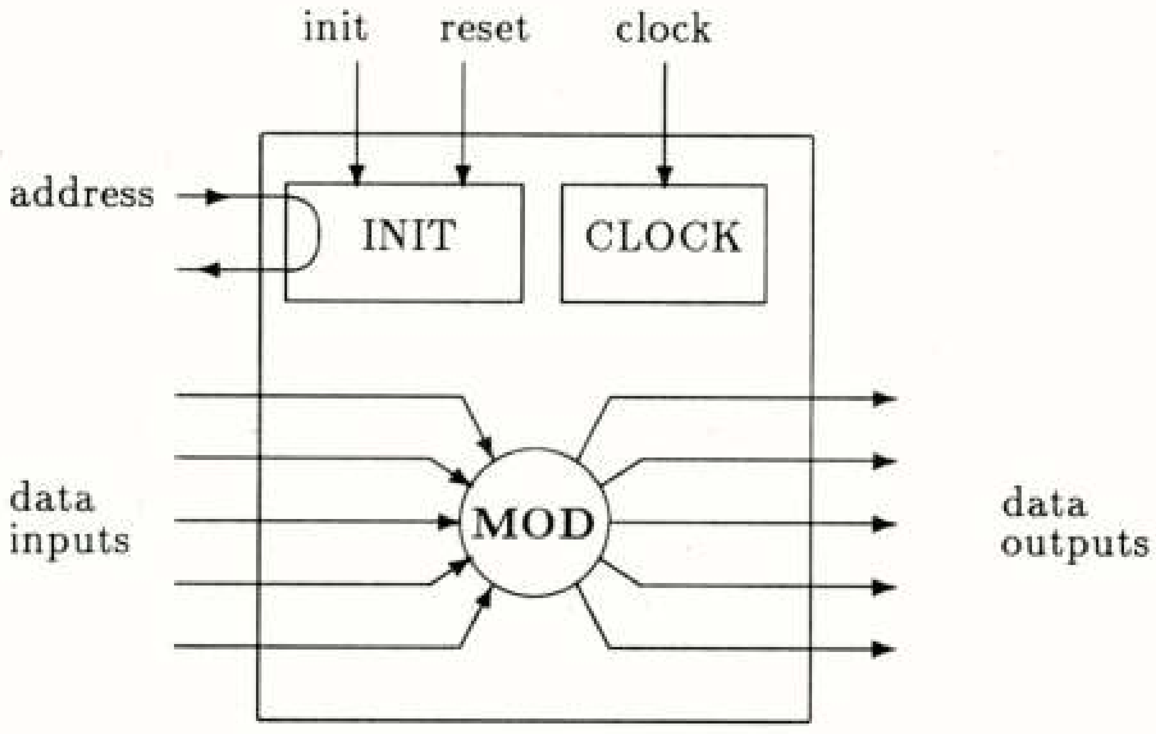


Figuur 3: Systeem, opgebouwd met VSP's

Alle type signalen van een willekeurig real-time videosysteem (dus luminantie, chrominantie, R,G,B, synchronisatiesignalen, control-signalen) met willekeurig *format* kunnen dus elk met hun eigen bemonsteringsfrequentie, in tijdmultiplex in eenzelfde VSP worden bewerkt; bovendien kunnen meerdere VSP's zonder communicatieoverhead samenwerken in een *data flow*-achtige structuur, zie Fig. 3, doordat we zijn uitgegaan van een synchrone communicatie die ook gestuurd wordt door de VSPs zelf: communicatie maakt deel uit van de programmering. Bovengenoemde punten leiden tot een IC met een complexe en niet voor de hand liggende processing en communicatie; voor de gebruiker wordt dit echter afgeschermd door de programmeertools.

VSP-device

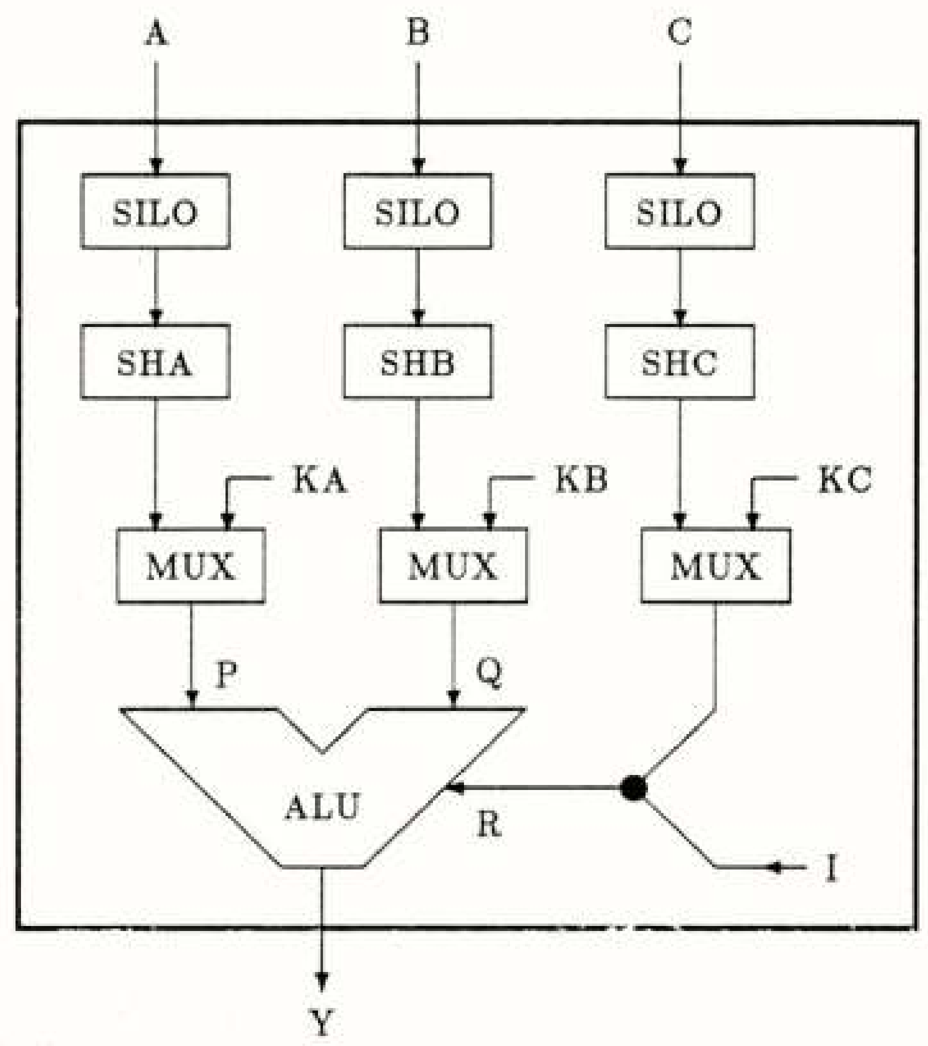
Fig. 4 toont de architectuur van de VSP op het hoogste niveau: een verwerkings-module 'mod' met 5 in- en 5 uitgangen, elk 12 bit parallel; een initialisatie-deel 'init' dat voor het laden van de programma's zorgt; en een klokcircuit 'clock' dat de interne kloksignalen verzorgt.



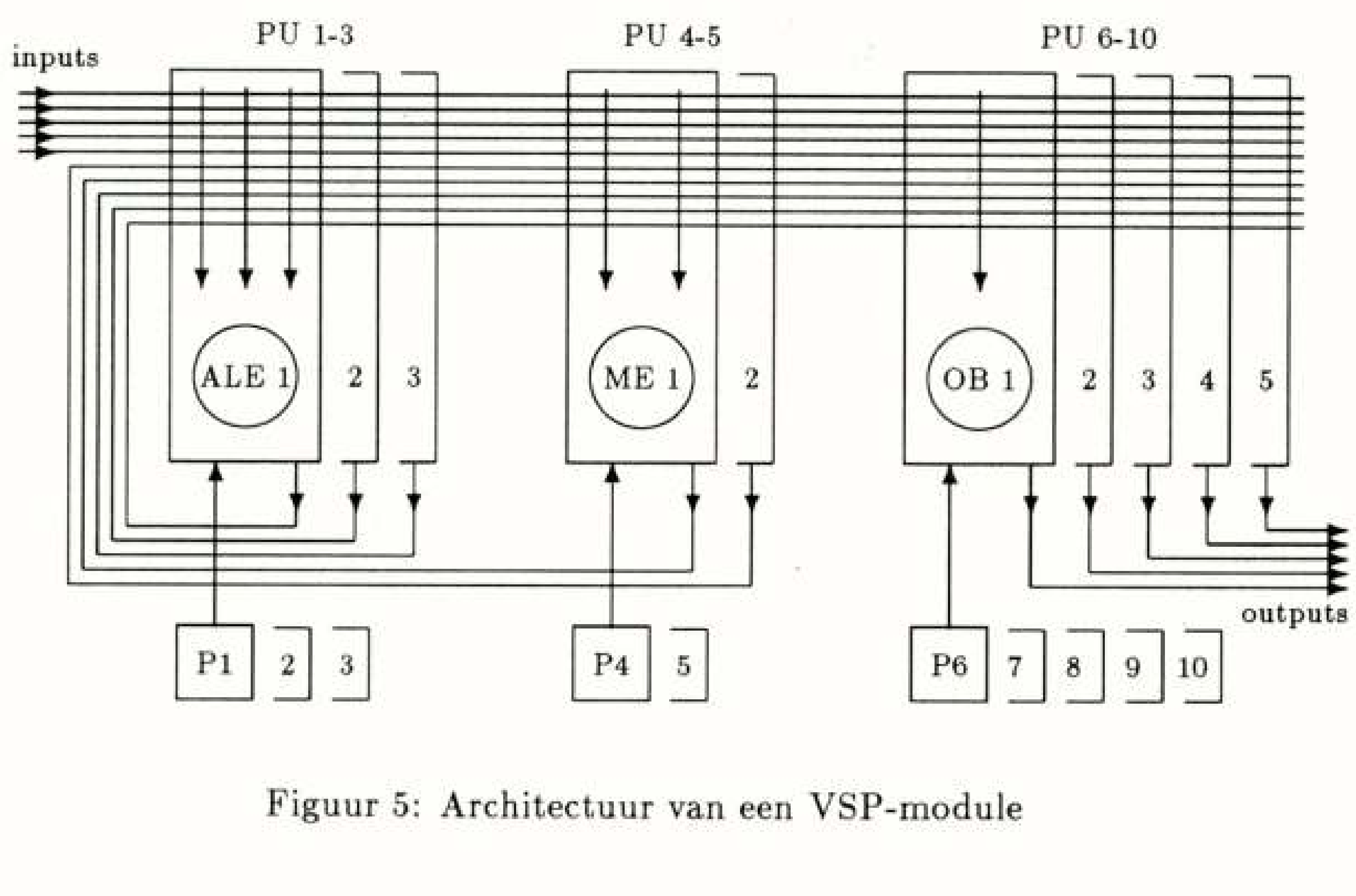
Figuur 4: De VSP intern

Fig. 5 toont de architectuur van de verwerkingseenheid 'mod': 10 onafhankelijke parallel opererende programmeerbare eenheden *programmable units*; 'PU 1-10'), elk met zijn eigen (deel)programmageheugen. De programma's bevatten maximaal 16 PU-instructies; het aantal wordt door de tools bepaald aan de hand van de door de gebruiker opgegeven herhalingsfrequenties bij de bewerkingen. De programma's worden cyclisch herhaald. Elke klokslag wordt uit elk programma een nieuwe instructie gestart. Zo'n PU-instructie bestaat uit een combinatie van instructies voor al de deelcomponenten uit de PU. Een gedistribueerde schakelmatrix zorgt voor een volledig flexibele communicatie, gestuurd vanuit de programmageheugens.

De eerste drie PU's bevatten naast (een deel van) de schakelmatrix een 'ALE' (*arithmetic and logic element*) bestaande uit een 3-input ALU, 3 multiplexers, 3 Barrel Shifters en 3 'Silo's', zie Fig. 6. De derde ALE-ingang heeft invloed op de instructie van de ALU en maakt zo data-afhankelijke bewerkingen en 3-operand-bewerkingen mogelijk (zoals data-afhankelijk schakelen en 3 bit partiële vermenigvuldigingen). Met de Barrel Shifters kunnen logische en aritmetische schuifoperaties worden uitgevoerd en met de Silo's (programmeerbare vertragingselementen) kunnen samples in de tijd worden verschoven, herhaald, of gepermuteerd, zowel ten behoeve van de bewerkingen als van de communicatie.

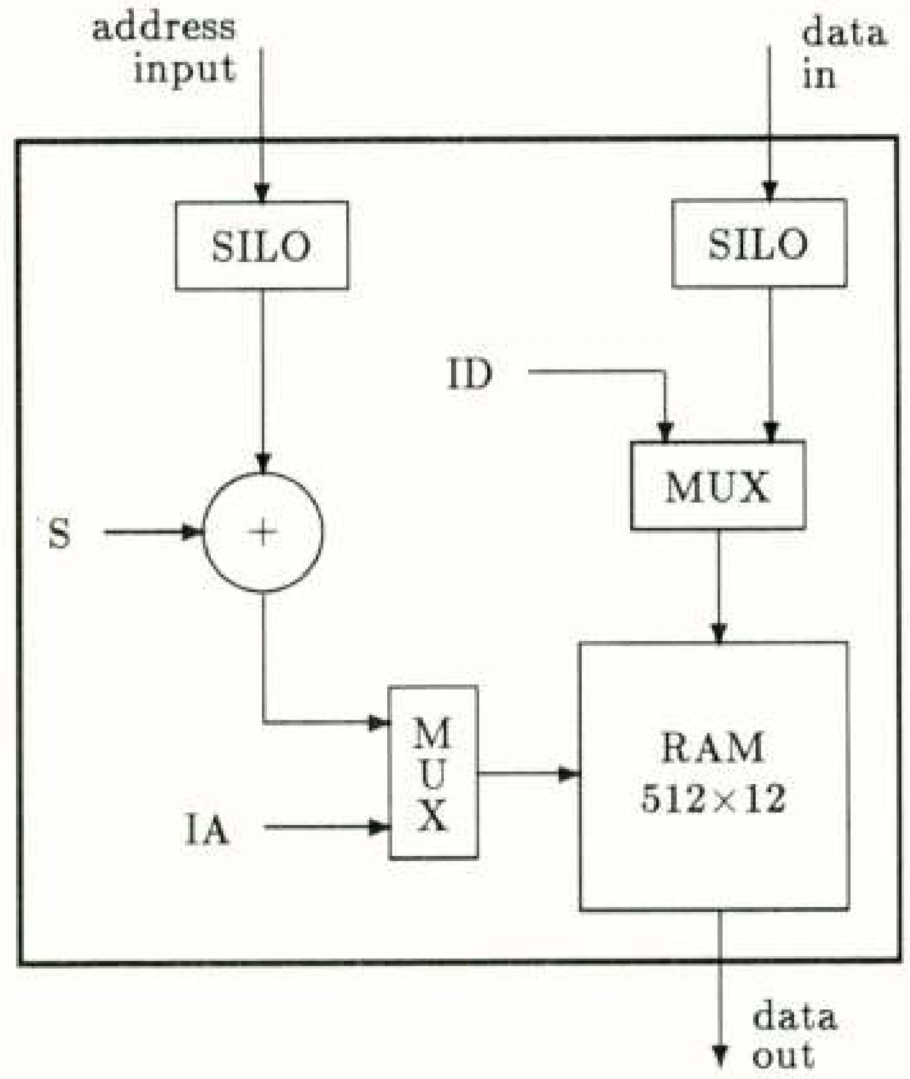


Figuur 6: Architectuur van de ALE-eenheid

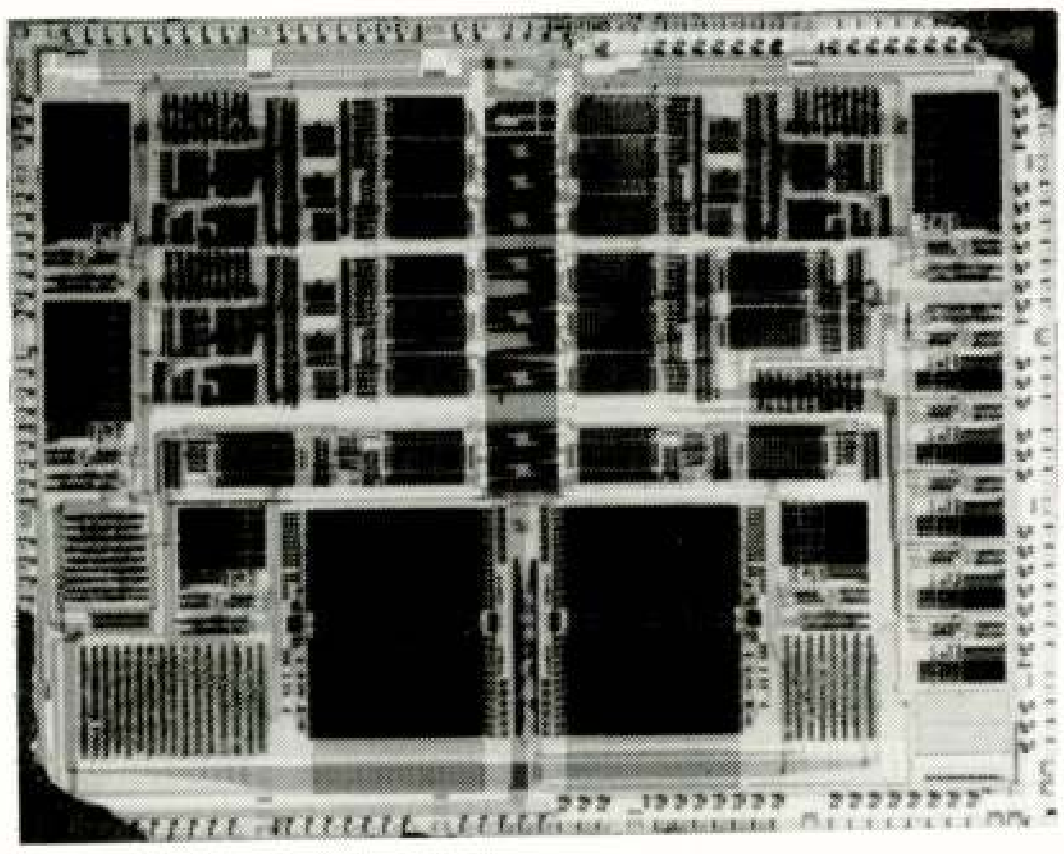


Figuur 5: Architectuur van een VSP-module

Twee andere PU's bevatten een *memory element* 'ME', see Fig. 7, met daarin een 512×12 bit SRAM. Data en adres komen van de schakelmatrix; elke ME-instructie bevat een constante die bij het adres wordt opgeteld. Met deze fijnadressering kunnen een aantal features worden gerealiseerd, zoals *paging*, waar hier niet verder op wordt ingegaan. Bij gebruik als *look-up table* moet het RAM vooraf worden geladen. Dat gebeurt via de ID en IA ingangen.



Figuur 7: Architectuur van de ME-eenheid



Figuur 8: Foto van het VSP-IC

IC-KARAKTERISTIEKEN	
chipgrootte	$120 mm^2$
technologie	1.6μm CMOS
# transistors	206k
klokfrequentie	27 MHz
instructiesnelheid	270 MIPS
rekenkracht	1400 MOPS
dissipatie	< 1.5 W
dataformaat	12 bits
behuizing	208 pens PGA
# gebruikte pennen	158
I/O-pennen	120

Tabel 1: IC-karakteristieken

De vijf overige PU's zijn *output buffers* 'OB', die de communicatie met de buitenwereld verzorgen en uit een silo en een uitgangs-stuurcircuit (met programmeerbare spanningsslag) bestaan. De communicatie tussen de VSP's onderling en de rest van het systeem zit dus opgesloten in de programma's van de OB's, terwijl de interne communicatie verdisconteerd zit in de programma's van de andere PU's.

De VSP is geïntegreerd; Fig. 8 toont een foto van het IC en de belangrijkste karakteristieken ervan zijn weergegeven in Tabel 1. De rekenkracht van de processor is uitgedrukt in het aantal instructies per seconde (MIPS), waarbij elke instructie weer bestaat uit een aantal deeloperaties (schakelfuncties, vertragingen, shifts, multiplexing, ALU-instructie, lees- en schrijfinstructie bij ME, constantes toevoegen) die elk op zich gekozen kunnen worden door de gebruiker. Als we al die operaties optellen, dan komen we tot het aantal MOPS dat in de tabel genoemd wordt. Omdat we het hier over combinaties van operaties hebben, moet de betekenis van het aantal MOPS in relatie tot de efficiëntie van het programma worden gezien.

Het programmeertrajekt bij de VSP

Het IC-ontwikkeltrajekt bij de VSP-aanpak heeft in handen gelegen van expert IC-ontwerpers die het complexe IC op een efficiënte wijze hebben gerealiseerd, door een optimale mix van de eerdergenoemde ontwerpmethodieken te gebruiken. Voor de uiteindelijke gebruiker is het IC-ontwerpen vervangen door het programmeren van de VSP. Een uitgebreide programmeeromgeving zorgt ervoor dat de complexiteit van het IC voor hem afgeschermd wordt en dat het programmeren een relatief eenvoudige taak is geworden.

In het kort komt het ontwerpen van een VSP-gebaseerd systeem neer op 4 stappen. De eerste stap bestaat uit het opbouwen van een functioneel schema op het niveau van de basisinstructies van de VSP (*signal flow graph*). De tweede uit het definiëren van het interconnectiepatroon tussen de te gebruiken VSPs en de in- en uitgangen. De derde stap bestaat uit het toewijzen van de basisoperaties (instructies) in het functionele schema aan de PU's in de verschillende VSP's (*assignment*). Hierbij kan optimaal gebruik gemaakt worden van de hardware door operaties op lagere snelheid dan de verwerkingssnelheid van de PU's (27 MHz) aan éénzelfde PU toe te kennen, totdat de maximale verwerkingssnelheid voor die PU is bereikt.

Tenslotte volgt de vierde stap: het toekennen van tijdsloten aan de verschillende operaties (*scheduling*). Dit alles gebeurt grafisch, interactief, en op een voor de ontwerper voor de hand liggende wijze.

Afhankelijk van de complexiteit van een algoritme neemt de programmering ervan een paar uur tot een paar dagen in beslag; aanpassingen aan een reeds afgebeeld algoritme (optimalisatie) nemen in de regel slechts een paar minuten tot een uur in beslag; tijden die dus verwaarloosbaar zijn als we ze vergelijken met de tijd die nodig is voor het ontwikkelen van een IC.

5 Samenvatting

Real-time video stelt hoge eisen t.a.v. rekenkracht en verwerking van hoge datastromen. Dat leidt tot problemen bij zowel de simulatie met computers als de hardware voor prototyping en uiteindelijke produktimplementatie. Er is daarom grote behoefte aan efficiënte IC-implementatiemethoden, die zowel kunnen voldoen aan de hoge IC-specificaties, als aan de eisen t.a.v. snelheid en flexibiliteit voor het ontwerptrajekt. Een goede afweging van de voor- en nadelen van de beschikbare methoden is daarom erg afhankelijk van de door de applikatie gestelde eisen. In dit artikel is op deze probleemstelling ingegaan.

Een specifieke aanpak, gebaseerd op een programmeerbare processor voor real-time video, is verder uitgewerkt. Deze methode vervangt het IC-ontwerptrajekt door een programmeertrajekt en leidt zo tot een snel ontwerptrajekt met korte iteraties en dus efficiënte verificatie en optimalisatie, tot gegarandeerd werkende hardware zonder de risico's en kosten van een IC-ontwikkeling en tot flexibiliteit en extra functionaliteit in de vorm van herprogrammeerbaarheid.

Veel ontwikkelingen op het gebied van IC-implementatie van algoritmes voor real-time video zijn momenteel gaande; de processoraanpak neemt hierbij een bijzondere plaats in.

References

[1] A.H.M. van Roermund, P.J. Snijder, H. Dijkstra, C.G. Hemeryck, C.M. Huizer, J.M.P. Schmitz, and R.J. Sluijter. A general-purpose programmable video signal processor. *IEEE Transactions on Consumer Electronics*, 35(3):, August 1989.

Voordracht gehouden tijdens de 374e werkvergadering.

Conferentieaankondigingen

IIT90. Ion Implantation Technology. University of Surrey 30th July to 3rd August 1990.
Contactadres: Mike Jackson II90 School. Department of Electronic and Electrical Engineering. University of Surrey, Guildford. Surrey GU2 5 XH U.K. Tel. (0483) 509133. Fax (0483) 34139.
Aan deze conferentie gaat van 26-28 juli 1990 een Technical School vooraf over "Science and Technology of Ion Implantation into semiconductors.

De cursus "Optica en lasers" verzorgd door de stichting Optel wordt najaar 1990 herhaald in Nijmegen.
Nadere informatie: Mevr. G. van Heugten, tel. 080 - 528800.

Pato.
Digitale signaalverwerking 12, 13, 19, 20, 26 en 27 november 1990 te Eindhoven.
Elektro-magnetische compatibiliteit 15, 16, 22, 23, 29 en 30 november 1990 te Eindhoven.
Nadere informatie: Bureau van het Pato-orgaan, Postbus 30424, 2500 GK 's-Gravenhage. Tel. 070 - 3644957.

Symposium 1990 over Optische transducenten.
15-16 November in Enschede, Universiteit Twente.
Contactadres: Mevr. R. Rietdijk, 3T, Postbus 545, 7500 AM Enschede. Tel. 053 - 336633.

Concept programma NERG 1990

WV + 381
25 september?
Bitstreaming and source coding
Nat. Lab.

WV + 382
(NERG + IEEE)
6 november
Satellietnavigatie
NLR?

WV + 383
30 november?
EMC (ESD, EMI)
RNL, Leidschendam

WV = werkvergadering, uitsluitend voor NERG-leden;
WV + = werkvergadering, voor NERG-, AES- en IEEE-leden;
WB = werkbezoek, uitsluitend voor NERG-leden;
JV = jaarvergadering, uitsluitend voor NERG-leden;
SMR = semiminar.

Tijdschrift van het Nederlands Elektronica- en Radiogenootschap

Inhoud

deel 55 – nr. 2 – 1990

Druk: de Witte, Eindhoven.